

Thu Mar 13 17:59:43 2003

us-09-509-595b-1.rn1

GenCore version 5.1.4.D5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: February 14, 2003, 12:23:35

(without alignments)
8796.662 Million cell updates/sec

Title: US-09-509-595B-1
Perfect score: 2245
Sequence: 1 cgggacacagccgagcg...

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_MA :
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq :
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq :
3: /cgn2_6/ptodata/1/ina/5A_COMB.seq :
4: /cgn2_6/ptodata/1/ina/5B_COMB.seq :
5: /cgn2_6/ptodata/1/ina/5A_COMB.seq :
6: /cgn2_6/ptodata/1/ina/5B_COMB.seq :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	150	6.7	162450	4	US-09-345-882-1
2	149.6	6.7	98844	4	US-09-791-211-10
3	148.8	6.6	4129	2	US-08-370-319C-12
4	148.8	6.6	4129	4	US-09-224-834-12
5	144.8	6.4	81001	4	US-08-687-080-111
6	144.2	6.4	2448	2	US-08-750-580-1
7	143.8	6.4	7680	4	US-09-210-748A-3
8	143.8	6.4	72928	3	US-09-009-913-1
9	143	6.4	99500	4	US-09-798-096-10
10	142.6	6.4	29629	4	US-09-341-587-3
11	142.2	6.3	28720	4	US-09-729-995-3
12	142.2	6.3	45716	4	US-08-965-048-5
13	142.2	6.3	45716	4	US-08-965-048-5
14	141.2	6.3	31571	1	US-08-323-443B-1
15	141.2	6.3	53526	3	US-08-658-136-2
16	141.2	6.3	53577	4	US-08-658-136-2
17	140.8	6.3	14796	4	US-08-975-080-35
18	140.8	6.3	14796	4	US-08-975-080-35
19	140.8	6.3	14796	4	US-08-975-080-35
20	140.2	6.2	15297	4	US-09-496-694B-3
21	139.8	6.2	45546	4	US-09-817-180-3
22	139.4	6.2	648	4	US-09-146-053-6
23	139.4	6.2	84495	4	US-09-385-982-332
24	138.6	6.2	16389	4	US-09-797-906-3
25	138.4	6.2	59065	4	US-09-741-154-3
26	138.4	6.2	59065	4	US-09-813-817-3
27	138.2	6.2	9365	4	US-09-978-197-3
					US-09-608-285A-8

28	138.2	6.2	9365	4	US-09-350-836B-8	Sequence 8, Appl1
29	138.2	6.2	9365	4	US-09-370-265-8	Sequence 8, Appl1
30	138.2	6.2	14747	4	US-09-608-285A-42	Sequence 42, Appl1
31	138.2	6.2	15977	4	US-09-608-285A-59	Sequence 59, Appl1
32	136.8	6.1	9734	4	US-09-347-114A-80	Sequence 80, Appl1
33	136.8	6.1	49136	4	US-09-422-869-1	Sequence 80, Appl1
34	136.6	6.1	5590	4	US-09-050-159-129	Sequence 129, Appl1
35	136.6	6.1	162450	4	US-09-345-882-1	Sequence 1, Appl1
36	136.2	6.1	11558	5	PCT-US93-06251-23	Sequence 129, Appl1
37	136.2	6.1	1701	4	US-09-078-294-9	Sequence 23, Appl1
38	136.2	6.1	17041	1	US-08-076-011-1	Sequence 23, Appl1
39	136.2	6.1	87350	3	US-08-781-891-79	Sequence 1, Appl1
40	135.8	6.0	36159	4	US-09-791-211-3	Sequence 79, Appl1
41	135.8	6.0	14581	4	US-09-749-588-3	Sequence 3, Appl1
42	135.6	6.0	22481	4	US-08-520-373D-4	Sequence 3, Appl1
43	135.6	6.0	22481	5	US-08-367-841A-43	Sequence 4, Appl1
44	135.6	6.0	22481	5	PCT-US95-07201-43	Sequence 43, Appl1
45	135.6	6.0	22484	4	US-09-875-223-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-09-345-882-1/c
Sequence 1, Application US/09345882
Patent No. 6399373
GENERAL INFORMATION:
APPLICANT: Bouquellet, Lydie
TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-
FILE REFERENCE: GENSET 031A
CURRENT APPLICATION NUMBER: US/09/345,882
PRIOR FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: US 60/091,315
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/111,909
NUMBER OF SEQ ID NOS: 140
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 162450
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 72794
OTHER INFORMATION: 5-124-273 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 88073
OTHER INFORMATION: 5-127-261 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 90842
OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 93714
OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
FEATURE:
NAME/KEY: allele
LOCATION: 97122
OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99098
OTHER INFORMATION: 5-130-257 : polymorphic base A or G

NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
FEATURE:
NAME/KEY: allele
LOCATION: 88050..88096
OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
FEATURE:
NAME/KEY: allele

LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
FEATURE:
NAME/KEY: allele
LOCATION: 90819..90865
OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177

us-09-509-595b-1.rni

Page 4

Query Match	6.6%	Score 148.8;	DB 2;	Length 4129;
Best Local Similarity	84.0%;	Pred. No. 1.8e-21;		
Matches 168; Conservative	0;	Mismatches 32;	Indels 0;	Gaps 0

QY	1997	TACTAAAAATATATAAATTAGCTGGCTGTGTGTGTGTGGTGGCTGTAAATCCAGCTACTG	2056
Db	2965	TACTAAAAATATACAAAAATTATAGCCACACCGTGGTGTGCATGCTGTAAATCCAGCTACTG	3024
QY	2057	GGACCTCGAGGACATGAGAACTACTTGAACTCGGAGGTGGAAGTTGCAGTACGTGAAAT	2116
Db	3025	GGACCTCGAGGACGAGAAATCACTTGAAACCGGAGAGTGTGCACTGTGCGCGAAT	3084
QY	2117	TGCGCCACTGCATCCACTGTGTGTGTGGCAAGATGTGAGACTCGGTCCAAATAACAAACA	2178
Db	3085	AGTGCACATGCATCCGACGCTGGAGACAGACAAACTCCATCTCCAAATAATAATANA	3144
QY	2177	AAAAACACATACATAAAT	2196
Db	3145	TAAATTAATTAATTAAGT	3164

RESULT 4
US-09-224-834-12
Sequence 12, Application US/09224834

GENERAL INFORMATION:
APPLICANT: Bichard, Vincent; Van Pel, Alne;
APPLICANT: Traversari, Catia; W Ifel, Thomas; Coulle, Pierre;
APPLICANT: Boon-talleur, Thierry; de Plaen, Etienne
TITLE OF INVENTION: ISOLATED NUCLEIC ACID SEQUENCE CODING FOR A
TITLE OF INVENTION: TUMOR REJECTION ANTIGEN PRECURSOR PROCESSED TO AT LEAST ONE T
TITLE OF INVENTION: REJECTION ANTIGEN PRESENTED BY HLA-A2

ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA

COMPUTER READABLE FORM: Diskette, 5.25 inch, 360 kb storage medium type.

COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS

CURRENT APPLICATION DATA:

FILING DATE:
 CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/370,319

FILED DATE: 06/03/07
PRIOR APPLICATION DATA: 06/03/07

FILING DATE: 18-MAR-1993

NAME: Hanson, No. 6201111man D.
REGISTRATION NUMBER: 30, 946

REFERENCE/DOCUMENT NUMBER:
TELECOMMUNICATION INFORMATION

TELEPHONE: (212) 838-3884
TELEFAX: (212) 838-3884

SEQUENCE CHARACTERISTICS:—
LENGTH. 4129 base pairs

TYPE: nucleic acid
STRANDEDNESS: double

FEATURE:

Query Match	6.6%;	Score 148.8;	DB 4;	Length 122;
Best Local Similarity	84.0%;	Pred. No. 1.8e-21;		
Matches 168;	Conservative	0;	Mismatches 32;	Indels 0;
			Gaps	0

[illegible]

RESULT 5
US-09-750-580-1/c
; Sequence 1, Application US/09750580
; Patent No. 6455280

APPLICANT: Yen, Frances
APPLICANT: Denison, Blake
APPLICANT: Bour, Barbara
APPLICANT: Bihain, Bernard
APPLICANT: Dumas Milne Edwards, Jean-Baptiste
APPLICANT: Duclert, Aymeric

APPLICANT: SALTER, CLO, DARRIS
TITLE OF INVENTION: METHODS AND COMPOSITIONS
FILE REFERENCE: 89, US2, CIP
CURRENT APPLICATION NUMBER: US 09/097,750, 588
CURRENT FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: US 09/599,362

PRIOR FILING DATE. 2000 06 21
PRIOR APPLICATION NUMBER: PCT/IB00/0101
2000 06 21

PRIOR APPLICATION NUMBER: PCT/IB99/02038
 PRIOR FILING DATE: 1999-12-20

PRIOR APPLICATION NUMBER: US 45/405/000
PRIOR FILING DATE: 1999-12-21

PRIOR APPLICATION NUMBER: 1998-12-22
PRIOR FILING DATE: 1998-12-22

PRIOR FILING DATE: 1999-06-25

```

; SOFTWARE: Patent.pm
; CEO ID NO 1

```

```
LENGTH: 81001  
TYPE: DNA
```

ORGANISM: MONO-CELLULAR
FEATURE: FEEDING

LOCATION: 10946..12946
 covered INFORMATION: 5' regulatory region

NAME/KEY: exon
LOCATION: 12947..12958

OTHER INFORMATION: CRO-
NAME/KEY: exon

OTHER INFORMATION: exon 2

	Query Match	6.4%	Score 144.8;	DB 4:	length 81001;
	Best Local Similarity	83.7%;	Pred. No. 1.9e-20;		
	Matches 164; Conservative	0;	Mismatches 32;	Indels 0;	Gaps 0.
QY	1396 CTACTAAAATATATAAATTAGCTGGTGTGTTGTTGGTGTCCTTAATCCACGACTACT				
Db	64921 CTTATTT				
QY	2056 GGAGAGCCTTAGCGATGAGATCATCTTGAACCTGGGAGGAGGAGTTGCAGTAGCTGAGA				
Db	64861 AGSAGGCTTAGCGACGAGAGATCACTTGAACC GGGAGGAGGAGTTGCAGTAGCTGAGA				
QY	2116 TTGCCGCACCTGCATCCACGTCTGGTGGCAGAGTGAGACTCCGCTCTCAAAAAAACAAACA				
Db	64801 TTGCCGCACCTGCATCCACGTCTGGTGGCAGAGTGAGACTCCGCTCTCTATCAAAAAAAAAAAC				
QY	2176 AAAAAACCACTATACA 2191				
Db	64741 AAAAAATTAATTAATA 64726				

RESULT 6
US-08-687-080-111
; Sequence 111, Application US/08687080
; Patent No. 5965427X h.u.v.v.
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolgoy
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof


```

QY 1075 CACCTGGCTGCTGCTCCCTCCGCTCCGGAGATCCCACTGGACCTGG 1125
      ||||||| ||| ||||||| ||| ||||||| ||||||| ||||
Db 492 CACCTGGCTGCTGCTCCCACTGTCAGAGATACCACTGGCTCTGG 542

RESULT 3
LOCUS A1510675 588 bp mRNA linear EST 12-MAR-1999
DEFINITION vx91b07.y1 Soares_thymus_2NBMT Mus musculus cDNA clone
ACCESSION IMAGE:1282549 5', mRNA sequence.
VERSION A1510675
KEYWORDS A1510675.1 GI:4409580
SOURCE EST.
ORGANISM house mouse.
           Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 588)
AUTHORS NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
        Contact: Robert Strausberg, Ph.D.
        Email: cgaps-remail.nih.gov
        This clone is available royalty-free through LNL; contact the
        IMAGE Consortium (info@image.lnl.gov) for further information.
        MGI:674349
        This read is a RESEQUENCE of a previously sequenced mouse clone
        correct orientation)
        Seq primer: -40RP from Gibco
        High quality sequence stop: 398.
        Location/Qualifiers
          1..588
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="IMAGE:1282549"
            /clone_1b="Soares_thymus_2NBMT"
            /sex="male"
            /tissue_type="Thymus"
            /dev_stage="4 weeks"
            /lab_host="DH10B"
            /note="Vector: pT73D-Pac (Pharmacia) with a modified
            polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
            was primed with a Not I - oligo(dT) primer [5',
            TGTACCAATGTCAGAGGAGCGCGCTTTTCTTTTCTTTTCTTTTCTTTT
            3']; double-stranded cDNA was ligated to Eco RI adaptors
            (Pharmacia), digested with Not I and cloned into the Not I
            and Eco RI sites of the modified pT73 vector. RNA
            provided by Dr. Bertrand Jordan. Library went through two
            rounds of normalization, and was constructed by Bento
            Soares and M. Fatima Bernaldo."
BASE COUNT 126 a 165 c 165 g 130 t 2 others
ORIGIN
Query Match 10.6%; Score 237.4; DB 9; Length 588;
Best Local Similarity 71.9%; Pred. No. 6,7e-30;
Matches 353; Conservative 0; Mismatches 128; Indels 10; Gaps 3;

```

```

QY 838 ----GGCAGTGGGAAGAACAAGGCCCGCAGCAGAGTGGCCCGAAGCCTCTGTCAGACC 894
      ||||| ||||||| ||||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 AGTGCATATTTGAAGAACAAGGCCCGCAGCAGAGTGGCCCGAAGCCTCTGTCAGACC 343
QY 895 AAGGAGACCCAGAGGAGCTGCGCCCGCGGAGAGTGAAGCTGAGGCGCAGCAGGAGC 954
      ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 344 AAGGAGACCCAGAGTCACTATACCTAGATGAGCAGAAAGTGGCCAGCAGATGTGGGG---- 399
QY 955 GTTCCGCGCCCTCTGCGCCCTCCCAAGTACCCAGCTCCACAGAAATAGAGAGAG 1014
      ||| ||||| ||||| ||||||| ||||| ||||| ||||| ||||| |||||
Db 400 --TCTTCCCTTCATCCCTTCCAGTAGAGCCCAAGCTTAACAGAAAGAGATGAGTG 457
QY 1015 TGTGCGCTGTGTGGGAGCGCGGAGACTATCTGCTGTACGCGCTG--CCCTGGGCGCTT 1073
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 458 TGGCGTGTGGCAGACCGGAGGTGAGCCTTATCTGTGTGAGCGGCTTNCCTCCGCGCTT 517
QY 1074 CCACCTGAGCCTGCTGCTGCTCCGCTCCGCGGAGAGATCCCACTGGAGCTGCTC 1133
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 518 CACCTGCTTGGCTGTCTCCACCTCTGCMNAGAACCCAGTGCCCTTTTGAGATTTGCT 577
QY 1134 CAGCTGCTGCTC 1144
      ||||| |||
Db 578 CTGCTTGTCTC 588

RESULT 4
LOCUS HSDML04M6 200 bp DNA linear GSS 17-AUG-1998
DEFINITION Homo sapiens DNA for trapped exon (ID MDL04M06), genomic survey
sequence.
ACCESSION AJ001114
VERSION AJ001114.1 GI:3426119
KEYWORDS GSS.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 200)
AUTHORS Scott,H.S.
TITLE Direct Submision
JOURNAL Submitted (14-AUG-1997) Scott H.S., Department of Genetics and
Microbiology, University of Geneva Medical School, 1 rue Michel
Servet, Geneva, 1211, SWITZERLAND
2 (bases 1 to 200)
REFERENCE Nagamine,K., Peterson,P., Scott,H.S., Kudoh,J., Minoshima,S.,
Heino,M., Krohn,K.J., Lalici,M.D., Mullis,P.E., Antonarakis,S.E.,
Kawasaki,K., Asakawa,S., Ito,F. and Shimizu,N.
TITLE Mutations in a novel zinc finger protein AIR are responsible for
autoimmune polyglandular disease type I (APECED)
JOURNAL Unpublished
FEATURES
source 1..200
Location/Qualifiers
  1..200
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="21"
    /map="q22.3"
    /note="trapped exon"
BASE COUNT 37 a 48 c 68 g 35 t 12 others
ORIGIN
exon
Query Match 7.6%; Score 170.8; DB 17; Length 200;
Best Local Similarity 90.7%; Pred. No. 1e-18;
Matches 175; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```

QY 893 CCAAGGAGCCAGGCGCGTGGAGTGAAGTGAAGTGGCTGGCGGACAGAGCGCA 952
 Db 121 CCAAGGAGCGCCAGGCGCGTGGAGTGAAGTGAAGTGGCTGGCGGACAGAGCGCA 180
 QY 953 GCGTTCCCGCCCC 965
 Db 181 GCGTTCCCTNNMC 193

RESULT 5
 LOCUS BE627856
 DEFINITION BE627856 451 bp mRNA linear EST 24-AUG-2000
 u049c07.y1 Soares_thymus_2NBMt Mus musculus cDNA clone
 IMAGE:3375276 5' similar to TR:Q920E3 Q920E3 AIRE PROTEIN. ; mRNA
 sequence.
 ACCESSION BE627856
 VERSION BE627856
 KEYWORDS BE627856.1 GI:9908250
 SOURCE EST.
 ORGANISM house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 451)
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:1084880
 Seg primer: -40RP from Gibco
 High quality sequence stop: 356.
 Location/Qualifiers
 1. 451
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:3375276"
 /clone_1lb="Soares_thymus_2NBMt"
 /sex="male"
 /tissue_type="thymus"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="vector: p77T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer 15,
 TGTACCAATCTGAAGTGGAGCGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGT
 3'; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p77T3 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through two
 rounds of normalization, and was constructed by Bento
 Soares and M. Fatima Bernaldo."

BASE COUNT 78 a 161 c 129 g 83 t
 ORIGIN

Query Match 7.2%; Score 162.6; DB 10; Length 451;
 Best Local Similarity 71.8%; Pred. No. 1.8e-17;
 Matches 245; Conservative 0; Mismatches 84; Indels 12; Gaps 2;

QY 1428 GTGCGAGATGTGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1487
 Db 1 GTGTGGGAGATGTGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 60
 QY 1488 CTGCGACTTCCAGCGGACCTCCGCGCGGAGCGGCGCTGGCTGAGATCTGTC 1547
 Db 61 CTGCGACTTCCAGCGGCGCGCGCGCGGAGCGGCGCTGGCTGAGATCTGTC 120
 QY 1548 AGGAGAGCTG-----ACCCAGCGCGCTGTGAGAGGAGTGTGCGCGCGCGCGCG 1601
 Db 121 TGGAGACTGT 180

QY 1602 CCTGGCCCTGTGGCGCTGCCAA-----GGATGACATGTGCGCATGACGAGCGCGCTGTGCA 1655
 Db 181 TCCAGCACCTGTGGCGCTGCCAAAGTAGGAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
 QY 1656 CAGGATGACCTGTGAGTCCCTTGTGAGCGAGACACCTTGTGATGATCTGTGAGTGGCG 1715
 Db 241 TAGGAGAGAGCTGTGAGTCCCTTGTGAGCGAGACTGTGATGAGGAGCATCTGTGAGTGGCG 300
 QY 1716 CATCCAGAGATGTGAGTCCCTTGTGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1756
 Db 301 CATCCAGAGATGTGAGTCCCTTGTGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 341

RESULT 6
 LOCUS AL036334/C
 DEFINITION DKE2P564P032.r1 564 (synonym: hfbz2) Homo sapiens cDNA clone
 DKE2P564P032 5', mRNA sequence.
 ACCESSION AL036334
 VERSION AL036334
 KEYWORDS AL036334.2 GI:5927737
 SOURCE EST.
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 653)
 Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann
 S. (Duesterhoeft, et al.)
 Unpublished (1999)
 On Jul 7, 1999 this sequence version replaced gi:5405917.
 CONTACT: Duesterhoeft A
 MIPS
 Am Klopferstr. 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 s1 sequence also available.
 This clone (DKE2P564P032) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
 source
 1. 653
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKE2P564P032"
 /clone_1lb="564 (synonym: 'hfbz2')"
 /tissue_type="brain"
 /dev_stage="fetal"
 /lab_host="X1-2Blue"
 /note="vector: pAMP1; Site_1: NotI; Site_2: SalI"

BASE COUNT 172 a 125 c 126 g 226 t
 ORIGIN

Query Match 6.9%; Score 154; DB 9; Length 653;
 Best Local Similarity 81.7%; Pred. No. 4.2e-16;
 Matches 178; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1997 TACTAAATAATATAATTAATGCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2056
 Db 224 TACTAAATAATATAATTAATGCTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 165
 QY 2057 GAGCCTGAGCGATGGAATCACTTGAATCGGAGGCTGTGAGTGTGAGTGTGAGAT 2116
 Db 164 GAGCCTGAGCGAGGAATCGCTTGCACCCAGAGGCGGAGGCTGTGAGTGTGAGAT 105
 QY 2117 TGGCGACATGTCCAGT 2176
 Db 104 CGGTCATTGTGCTTCCAGCTTGTGAGCGAGAGAGTGAAGCCATCTCAAAAAA 45

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 546)	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	Unpublished (1997)
		Contact: Robert Strausberg, Ph.D.	Email: cgapbs-remail.nih.gov	Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmerl-Buck, M.D., Ph.D.
		CNA Library Preparation: Life Technologies, Inc.	DNA Sequencing by: Washington University Genome Sequencing Center	clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
		Insert Length: 748	Std Error: 0.00	Seq primer: -40UP from Gibco
		High quality sequence stop: 427.	Location/Qualifiers	
		1. 546	/organism="Homo sapiens"	
		/db_xref="taxon:9606"	/clone="IMAGE:2442244"	
		/clone_lib="NCI-CGAP-Ut2"	/tissue_type="moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors"	
		/lab_host="DH10B"	/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: Salt; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.85 kb. Life Technologies catalog #:	
		11359-012"		
BASE COUNT	180 a	125 c	122 g	117 t 2 others
ORIGIN				
Query Match	6.8%	Score 151.6;	DB 9;	Length 546;
Best Local Similarity	84.5%	Pred. No. 1.1e-15;		
Matches 169;	Conservative 0;	Mismatches 31;	Indels 0;	Gaps 0;
QY	1996	CTACTAAATAATATATAAATATAGTGTGGTGTGGTGGCTGTGAATCCAGCTACAT	2055	
Db	303	CTACTAAAGATCAAAAATATAGTGTGGTGTGGTGGCTGTGATCTCCAGCTACATC	362	
QY	2056	GGAGGCTGAGGCGATGAGATCTCACTTGAATCTCGGGAGGTGAGGTTGCAGTACGTGAGA	2115	
Db	363	AGGAGGCTGAGGCGAGGAGAACTGCTTGAACCTGNGAGGCGGAGTTGCAGTACGTGAGA	422	
QY	2116	TTGCGGCACATCGCACTCGAGTCTGGTGGCGCAGAGATGAGAACTCGCTCAAAAAACAACA	2175	
Db	423	TTTCGCCACTGCTCCAGCTGGGCGACAGAGCAAGCAATCTCATCTCANAAAAA	482	
QY	2176	AAAAAACCATATACATAAA	2195	
Db	483	CAAGAACAACAAAAA	502	
RESULT 10				
AG147516/c				
LOCUS	AG147516	669 bp	DNA	linear
DEFINITION	Pan troglodytes DNA, clone: RP43-009H02.TJ, genomic survey			GSS 08-JAN-2002
ACCESSION	AG147516			
VERSION	AG147516.1	GI:16677194		
KEYWORDS	GSS.			
SOURCE	Pan troglodytes male lymphocytes DNA, clone: RP43-009H02.TJ.			
ORGANISM	Male BAC library clone:RP43-009H02.TJ.			
REFERENCE				
AUTHORS	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,			
	Totoki,Y., Watanabe,H. and Sakaki,Y.			

TITLE	BAC end sequences of Library RPCI-43
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 669) Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoi, Y., Watanabe, H. and Sakaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (02-NOV-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suenho-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/), Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT	Clones are derived from the chimpanzee BAC library RPCI-43 This BAC end was generated during the Red process and may have higher chance of clone tracking errors.
PRIMERS	
Sequencing: TJ	
LIBRARY	
Vector	: pBAC3.6
R.Site 1:	: EcoRI
R.Site 2:	: EcoRI.
Location/Qualifiers	
1..669	
/organism="Pan troglodytes"	
/db_xref="taxon:9598"	
/clone="Rp43-009H02.TJ"	
/sex="male"	
/cell_type="lymphocytes"	
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"	
BASE COUNT	173 a 168 c 123 g 204 t 1 others
ORIGIN	
Query Match	6.7%: Score 151.4; DB 17; Length 669;
Best Local Similarity	79.6%: Pred. No. 1.1e-15;
Matches 179: Conservative	0; Mismatches 46; Indels 0; Gaps 0;
QY 1973	GGAGGCTCCCTGGCTGTGACCTACATATAATTTAAATAATTACCTGGTGTGCTG 2032
Db 278	GGAGGACCCCATCTCTGCTATAAAAAAAAAAAAAAAAAATTACCGAGCTGTGTGCA 219
QY 2033	GGTGGCTGTAAATCCAGCATGATGGAGCCTGAGCATGAGATACACTTGAATCGGAG 2092
Db 218	CATGGCTGTAAATCCAGCATCTGTGGAGAGCTGAGCAGAGAAATCGCTTGAATCCAGAG 159
QY 2093	GTGAGGTTGACGATGAGTGAATTCGCCCATCTGCATCCAGTCTGTGCGCAGAGTGA 2152
Db 158	GCGGAGTTTGGCGGTGAGCGGAGATGGCAGCATCGACATCCAGCGTGGGTGCAAGAGTGA 99
QY 2153	GACCTCGTCTCAAAAACAAAACAAAACCAATACATTAATT 2197
Db 98	AACCTCTGTCTCAAAAAAAAAAAAAAAAAAAGATTAATT 54
RESULT 11	
AJ2520846/c	
ACCESSION	AJ2520846
DEFINITION	RPCI-11-23864.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-23864, DNA sequence.
VERSION	AJ2520846
KEYWORDS	AJ2520846.1 GI:10833356
SOURCE	GSS.
ORGANISM	human.
REFERENCE	Homo sapiens
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi; Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
REFERENCE	1 (bases 1 to 749)
REFERENCE	Zhao, S., Adams, M. D., Nierman, W., Malek, J., de Jong, P. and Venter
REFERENCE	J.C.
REFERENCE	BAC end sequences of library RPCI-11
REFERENCE	Unpublished (1997)
REFERENCE	Other GSSs: RPCI-11-23864.TVC
REFERENCE	Contact: Shaying Zhao
REFERENCE	Department of Eukaryotic Genomics
REFERENCE	The Institute for Genomic Research

High quality sequence stop: 450.

FEATURES	Location/Qualifiers
source	1. .606

BASE COUNT
ORIGIN

197 a 142 c 130 g 137 t

Query Match	6.7%	Score 149.4	DB 14	Length 606
Best Local Similarity	84.4%	Pred. No. 2.5e-15		
Matches 168	Conservative 0	Mismatches 31	Indels 0	Gaps 0

OY	1996	CTACTAAAAATATAAAAATTAGCTGGGTGNGNGTGGGGGCGCTGATATCCCACTACAT	20533
Db	305	CTACTAAAAAGTACAAAAATTAGCTGGGTGNGNGTGGGGGCGCTGATATCCCACTACTC	364
OY	2056	GGGAGCCTGAGGCATGAGATATCACTTGAACTCGGAGAGTGGAGGTTGCAGTGAGCTGAGA	2115
Db	365	AGGAGGCTGAGGCGAGGAGAACTCTTGAACTCGGAGAGCGGAGTTGCAGTGAGCTGAGA	424
OY	2116	TTGGCGCACTGCATCTCCATCTCTGTGTGGCAGAGTGTGAGACTCCGTCCTCAAAAACAAAAACA	2175
Db	425	TTGGCGCACTGCATCTCCACCTGTGGCGCAGAAAGAGCAAGATCTCATCTCAAAAAAAAAC	484
OY	2176	AAAAAACCCATATACATAA	2194
Db	485	AAAAGACAAAAAACAAAAA	503

Search completed: February 14, 2003, 16:46:08
Job time : 2892.11 secs

JOURNAL Patent: WO 918197-A 1 15-APR-1999;
 MAX PLANCK GESELLSCHAFT (DE); YASPO MARIE LAURE (DE)
 FEATURES Location/Qualifiers
 source 1..2245
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 419 a 765 c 697 g 364 t
 ORIGIN

Query Match 100.0%; Score 2245; DB 6; Length 2245;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGCGACACGCGCGCGGAGGCGCCACAGCCCGCGGAGCCCGGAGGCGCAAGGAGG 60
 Db 1 CGGGCGACACGCGCGCGGAGGCGCCACAGCCCGCGGAGCCCGGAGGCGCAAGGAGG 60
 QY 61 GCTGCGAGTGTCCCGGAGCCACCGCGCTCCGCCCGCGGAGTCCCGCGGAGCCCGC 120
 Db 61 GCTGCGAGTGTCCCGGAGCCACCGCGCTCCGCCCGCGGAGTCCCGCGGAGCCCGC 120
 QY 121 ATGCGCAGCGAGCGCGCGCTACGCGCGCTTGTAGGCTGCACCGCAGAGATCGGGT 180
 Db 121 ATGCGCAGCGAGCGCGCGCTACGCGCGCTTGTAGGCTGCACCGCAGAGATCGGGT 180
 QY 181 GCGGTGACAGCGCGCTTCCCACTGTCAGCGCGCTGTGACAGAGAGTGTCCCGAG 240
 Db 181 GCGGTGACAGCGCGCTTCCCACTGTCAGCGCGCTGTGACAGAGAGTGTCCCGAG 240
 QY 241 GACAAAGTTTACAGAGAGCGCTTCACTGTAAGAAAAAGAGGCGTCCCGCAGGCTTCCAC 300
 Db 241 GACAAAGTTTACAGAGAGCGCTTCACTGTAAGAAAAAGAGGCGTCCCGCAGGCTTCCAC 300
 QY 301 GCGCTCTGTCTGTGCTGTGCTGACCCAGAGCTTCACAGCCATCTGAGACTTGTGAGGGTG 360
 Db 301 GCGCTCTGTCTGTGCTGTGCTGACCCAGAGCTTCACAGCCATCTGAGACTTGTGAGGGTG 360
 QY 361 CTGTTCAAGAGCTCAACCTGAGAGCGGTATGGCGGCGTGCAGCCCATCTGAGACAGCTTC 420
 Db 361 CTGTTCAAGAGCTCAACCTGAGAGCGGTATGGCGGCGTGCAGCCCATCTGAGACAGCTTC 420
 QY 421 CCCAAAGATGTGACCTCAGCCAGCCCGGAAAGGAGGAGGAGCCCGCGCTGCCAAG 480
 Db 421 CCCAAAGATGTGACCTCAGCCAGCCCGGAAAGGAGGAGGAGCCCGCGCTGCCAAG 480
 QY 481 GCTTTGGTACCAGCCCACTCCCAAGAGAGAGGCTTGAAGAGAGCTGAGCT 540
 Db 481 GCTTTGGTACCAGCCCACTCCCAAGAGAGAGGCTTGAAGAGAGCTGAGCT 540
 QY 541 GCGCGCCACAGCCCTGATCTCCAAAGGGGACCGCCAGCCCATGCTCAACTGAAGGCC 600
 Db 541 GCGCGCCACAGCCCTGATCTCCAAAGGGGACCGCCAGCCCATGCTCAACTGAAGGCC 600
 QY 601 AAGCCCCCAAGAGAGCGGAGAGAGCGAGAGAGCGGCTTCCACTGAGGAGCGG 660
 Db 601 AAGCCCCCAAGAGAGCGGAGAGAGCGAGAGAGCGGCTTCCACTGAGGAGCGG 660
 QY 661 ATTCAACACATGTACGTTAGTCCAGAGAGCTGTGCCATGTCTCCGAGGAGAGTCCCG 720
 Db 661 ATTCAACACATGTACGTTAGTCCAGAGAGCTGTGCCATGTCTCCGAGGAGAGTCCCG 720
 QY 721 GGAGCCGAGGAGCGGCTGAGGGGATCTCATCAGCAGAGGTTTGAAGTCAAGGGGCTCC 780
 Db 721 GGAGCCGAGGAGCGGCTGAGGGGATCTCATCAGCAGAGGTTTGAAGTCAAGGGGCTCC 780
 QY 781 AAGAAATGATCCAGGTTGGTGGGAGTTTACACTTCCAGCAAGTTTGAAGATCTCCGCG 840
 Db 781 AAGAAATGATCCAGGTTGGTGGGAGTTTACACTTCCAGCAAGTTTGAAGATCTCCGCG 840
 QY 841 AGTGGGAAGAACAGGCGCGCAGAGAGAGTGGCCGGAAGGCTTGTGTTCAAGCCAAAGGA 900
 Db 841 AGTGGGAAGAACAGGCGCGCAGAGAGAGTGGCCGGAAGGCTTGTGTTCAAGCCAAAGGA 900

QY 901 GCCCAGGCGCTGCCCGCGGTGAGGTAGGCTAGGCTGAGGCGCAGAGGAGCGCTTCCC 960
 Db 901 GCCCAGGCGCTGCCCGCGGTGAGGTAGGCTAGGCTGAGGCGCAGAGGAGCGCTTCCC 960
 QY 961 GCCCCTTGGCCCTCCCGCAGAGAGAGCCAGCTCCACAGAGAGAGAGAGAGAGAGTCCC 1020
 Db 961 GCCCCTTGGCCCTCCCGCAGAGAGAGCCAGCTCCACAGAGAGAGAGAGAGAGAGTCCC 1020
 QY 1021 GTGTGTGAGAGCGCGGAGAGCTCATGTGTGAGAGGCTCCCTCGGAGCTTCCACCTG 1080
 Db 1021 GTGTGTGAGAGCGCGGAGAGCTCATGTGTGAGAGGCTCCCTCGGAGCTTCCACCTG 1080
 QY 1081 GCGTGTCTGTCCCTCCGCTCCGAGAGATCCCAAGTGGAGACTGAGAGTCTCCAGCTGC 1140
 Db 1081 GCGTGTCTGTCCCTCCGCTCCGAGAGATCCCAAGTGGAGACTGAGAGTCTCCAGCTGC 1140
 QY 1141 CTGAGGCAACACTCCAGAGAGTGCACGCGCGGCGCAGAGAGCGCGCGCCAGAGGCA 1200
 Db 1141 CTGAGGCAACACTCCAGAGAGTGCACGCGCGGCGCAGAGAGCGCGCGCCAGAGGCA 1200
 QY 1201 CCGGTGAGACCCCGCTCCCGCGGCGCTTATGTCGCGGAGAGAGTAAAGAGTCCA 1260
 Db 1201 CCGGTGAGACCCCGCTCCCGCGGCGCTTATGTCGCGGAGAGAGTAAAGAGTCCA 1260
 QY 1261 CTTGGGGAACCCCTAGCGCGCATGAGACAGACTTCTTCAAGACACTGCGGCTCCG 1320
 Db 1261 CTTGGGGAACCCCTAGCGCGCATGAGACAGACTTCTTCAAGACACTGCGGCTCCG 1320
 QY 1321 CTTTTCAGAGCCCGCGGCGCAGAGGCTGTGACCTCCGCGCGCTTCACTGTGTG 1380
 Db 1321 CTTTTCAGAGCCCGCGGCGCAGAGGCTGTGACCTCCGCGCGCTTCACTGTGTG 1380
 QY 1381 GGTCTGAGGAGTACAGAGAACTGTGCTCTGTGTCGCGGCTGTGCGGAGATGT 1440
 Db 1381 GGTCTGAGGAGTACAGAGAACTGTGCTCTGTGTCGCGGCTGTGCGGAGATGT 1440
 QY 1441 AGGAGAGTGTGCGGCTGTACTACTGCGCGCTGTCTTCCACTGTGCGCTGCCACTTCCA 1500
 Db 1441 AGGAGAGTGTGCGGCTGTACTACTGCGCGCTGTCTTCCACTGTGCGCTGCCACTTCCA 1500
 QY 1501 GCGGACACCTCCCGCGCGGAGCGGCGCTGCGTGCAGATCTGCTAGAGAGAGTGAAC 1560
 Db 1501 GCGGACACCTCCCGCGCGGAGCGGCGCTGCGTGCAGATCTGCTAGAGAGAGTGAAC 1560
 QY 1561 CGAGCCCTGTGAGAGGAGGTGCTGAGCCAGCCCGCGCGCTTGGGCGCTGCC 1620
 Db 1561 CGAGCCCTGTGAGAGGAGGTGCTGAGCCAGCCCGCGCGCTTGGGCGCTGCC 1620
 QY 1621 AAGATGACACTGTCCAGTACAGAGCCCGCTGTGACAGAGGATACCTGGAGTCCCTTCTG 1680
 Db 1621 AAGATGACACTGTCCAGTACAGAGCCCGCTGTGACAGAGGATACCTGGAGTCCCTTCTG 1680
 QY 1681 AGCAGACACCTCTGATGGATCTGCACTGAGTGGGAGCATCAGAGAGTGGCGGCGG 1740
 Db 1681 AGCAGACACCTCTGATGGATCTGCACTGAGTGGGAGCATCAGAGAGTGGCGGCGG 1740
 QY 1741 GCCCCTTCCCTCTCTGACCCAGATGCGCGGAGATGACGCTGTATGAGAGAGTGTG 1800
 Db 1741 GCCCCTTCCCTCTCTGACCCAGATGCGCGGAGATGACGCTGTATGAGAGAGTGTG 1800
 QY 1801 AGAAGACACCTCTCTCTGACCTGTGGAAGCCGCGCGCTGGAGATCAAGAGGAGACAG 1860
 Db 1801 AGAAGACACCTCTCTCTGACCTGTGGAAGCCGCGCGCTGGAGATCAAGAGGAGACAG 1860
 QY 1861 CGCCACCTCTGTGATGATGTGCGGTGAAGAGCTGTGTTTCTGGGGAGACACAGCCAT 1920
 Db 1861 CGCCACCTCTGTGATGATGTGCGGTGAAGAGCTGTGTTTCTGGGGAGACACAGCCAT 1920
 QY 1921 CATGTGCTGGAATTAACCTTGCACCTCTCTACTCTGTGAAGTCCCGGAGAGCTTC 1980
 Db 1921 CATGTGCTGGAATTAACCTTGCACCTCTCTACTCTGTGAAGTCCCGGAGAGCTTC 1980
 QY 1981 TCTTGTGCTGTGAGCTACTATAAATAATATAAATAATAGTGGGTGTGCTGGTGCCTG 2040

AAHMHCFPAAGTSPGTGLRCSCSGVTPAPVEGLAPSPARLAPGPAKDDTASHE
PALHRLDLESLSHSTHTFDGILLQMAIOMARAPAPPSS"
polyA_signal
2013. . 2018
/gene="APECED"
BASE COUNT 357 a 716 c 644 g 310 t
ORIGIN

Query Match 89.9%; Score 2018.4; DB 9; Length 2027;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2019; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CGGGGCGACACCGCGGGGAGGGGCCCCACAGCCCCCGGGAGCCGAGCCCAAGCAGGG 60
DB 8 CGGGGCGACACCGCGGGGAGGGGCCCCACAGCCCCCGGGAGCCGAGCCCAAGCAGGG 67
QY 61 GCTGCGACAGTGTCCGGGGAGCCACCGCGTCCGCCAGCCCGGGTCCCGGCGCCACCCC 120
DB 68 GCTGCGACAGTGTCCGGGGAGCCACCGCGTCCGCCAGCCCGGGTCCCGGCGCCACCCC 127
QY 121 ATGGCGAGGAGCGGGCGCTACGCGGCTTGTAGAGCTGACCGACGAGATCCGGGTG 180
DB 128 ATGGCGAGGAGCGGGCGCTACGCGGCTTGTAGAGCTGACCGACGAGATCCGGGTG 187
QY 181 GCGGTGAGACGCGCTTCCCACTGCTGACGCGGTGCTGACACAGAGCTGTCCCGAG 240
DB 188 GCGGTGAGACGCGCTTCCCACTGCTGACGCGGTGCTGACACAGAGCTGTCCCGAG 247
QY 241 GACAAGTTTTCAGAGACGCTTCAATCTGAAGAAAGAGGGGTGCCCCAGGCTTCCAC 300
DB 248 GACAAGTTTTCAGAGAGCGCTTCAATCTGAAGAAAGAGGGGTGCCCCAGGCTTCCAC 307
QY 301 GCCCTCTGCTCTGGCTGCTACCCAGAGCTCCAGCCATCTGACTTTCGAGGGTG 360
DB 308 GCCCTCTGCTCTGGCTGCTACCCAGAGCTCCAGCCATCTGACTTTCGAGGGTG 367
QY 361 CTGTTCAGAGACTCAACCTGAGAGGCTATGCGCGGTGACGCCATCTCTGAGAGCTTC 420
DB 368 CTGTTCAGAGACTCAACCTGAGAGGCTATGCGCGGTGACGCCATCTCTGAGAGCTTC 427
QY 421 CCCAAGATGTGACCTCAGCCAGCCCGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
DB 428 CCCAAGATGTGACCTCAGCCAGCCCGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 487
QY 481 GCTTTGTACCGCCACCCAGACTCCCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540
DB 488 GCTTTGTACCGCCACCCAGACTCCCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 547
QY 541 GCGCGCGACAGCGCTGACTCCAAAGGGGACCCCGACCCAGAGCTCTCACTGAAGGCC 600
DB 548 GCGCGCGACAGCGCTGACTCCAAAGGGGACCCCGACCCAGAGCTCTCACTGAAGGCC 607
QY 601 AAGCCCCCAAGAAAGCGGAGAGCGGACAGAGCAGAGCGCTTCCACTCGGGAGAGGG 660
DB 608 AAGCCCCCAAGAAAGCGGAGAGCGGACAGAGCAGAGCGCTTCCACTCGGGAGAGGG 667
QY 661 ATTCAACACCTGTGACGTTTCACTTCAAGAGCTGTGCGCATGTCTCCGGGAGCTCCG 720
DB 668 ATTCAACACCTGTGACGTTTCACTTCAAGAGCTGTGCGCATGTCTCCGGGAGCTCCG 727
QY 721 GAGAGCCGAGAGGCGCGTGAAGGGGATCTCATCCAGAGGTTTGAATCAGGGGCTCC 780
DB 728 GAGAGCCGAGAGGCGCGTGAAGGGGATCTCATCCAGAGGTTTGAATCAGGGGCTCC 787
QY 781 AAGAGTGCATCCAGGTTGTGGGAGTTCTACACTCCAGCAAGTTTGAAGACTCCGGC 840
DB 788 AAGAGTGCATCCAGGTTGTGGGAGTTCTACACTCCAGCAAGTTTGAAGACTCCGGC 847
QY 841 AGTGGGAAGAAAGAGCGCCGACAGAGAGTGGCCCAAGCTCTGTTCGAGCAAGGGA 900
DB 848 AGTGGGAAGAAAGAGCGCCGACAGAGAGTGGCCCAAGCTCTGTTCGAGCAAGGGA 907
QY 901 GCCCAGGCGCTGCCCGGTGAGGTAGAGCTGAGGCGCAGGAGGAGGAGGAGGAGGAGG 960
DB 908 GCCCAGGCGCTGCCCGGTGAGGTAGAGCTGAGGCGCAGGAGGAGGAGGAGGAGGAGG 967

QY 961 GCCCTCTGCGGCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1020
DB 968 GCCCTCTGCGGCTTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1027
QY 1021 GTGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1080
DB 1028 GTGTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1087
QY 1081 GCTGCTGTCCCTCCGCTCCGGGAGATCCCACTGTGGAGCTGAGAGTGTCCAGCTTC 1140
DB 1088 GCTGCTGTCCCTCCGCTCCGGGAGATCCCACTGTGGAGCTGAGAGTGTCCAGCTTC 1147
QY 1141 CTGAGGCAACAGTCCAGAGAGTGCAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1148 CTGAGGCAACAGTCCAGAGAGTGCAGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1207
QY 1201 CCGGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
DB 1208 CCGGTGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1267
QY 1261 CTGTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
DB 1268 CTGTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1327
QY 1321 CTTTCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
DB 1328 CTTTCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1387
QY 1381 GGTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
DB 1388 GGTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1447
QY 1441 ACGGAGTGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
DB 1448 ACGGAGTGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1507
QY 1501 GCGGAGACCTCCCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
DB 1508 GCGGAGACCTCCCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1567
QY 1561 CCAGCCCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
DB 1568 CCAGCCCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1627
QY 1621 AAGGATGACACTGCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680
DB 1628 AAGGATGACACTGCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1687
QY 1681 AGGAGACACACTTGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1740
DB 1688 AGGAGACACACTTGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1747
QY 1741 GCCCCTTCCCTCTGAGCCCGAGATGGCGGGGACATGAGAGCTGTGATGAGAGAGTCTG 1800
DB 1748 GCCCCTTCCCTCTGAGCCCGAGATGGCGGGGACATGAGAGCTGTGATGAGAGAGTCTG 1807
QY 1801 AGAAGGACACTTCTTCTGAGTCTGAGAGGCGGCGGCTGGGATCAAAAGAGGAGAG 1860
DB 1808 AGAAGGACACTTCTTCTGAGTCTGAGAGGCGGCGGCTGGGATCAAAAGAGGAGAG 1867
QY 1861 GCGCACCTTGTGACAGTGTGCTGAGGAGTGAACAGCTGTGTGTTCTGGGAGACACAGCAT 1920
DB 1868 GCGCACCTTGTGACAGTGTGCTGAGGAGTGAACAGCTGTGTGTTCTGGGAGACACAGCAT 1927
QY 1921 CATGTGCTGAGAAATTAACCCGACCTTCTACTCTGAGAGTCCCGGAGAGCTTC 1980
DB 1928 CATGTGCTGAGAAATTAACCCGACCTTCTACTCTGAGAGTCCCGGAGAGCTTC 1987
QY 1981 TCTTGTGCTGTGACCTACTAAATAATATATATATATATATATATATATATATAT 2020
DB 1988 TCTTGTGCTGTGACCTACTAAATAATATATATATATATATATATATATATATAT 2027

REFERENCE Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 1 (bases 1 to 1656)
 AUTHORS Shi,J.D., Wang,C.Y., Marron,M.P., Ruan,Q.G., Huang,Y.Q.,
 Dettler,J.C. and She,J.X.
 TITLE Chromosomal localization and complete genomic sequence of the
 murine autoimmune regulator gene (Aire)
 JOURNAL Autoimmunity 31 (1), 47-53 (1999)
 MEDLINE 20059142
 PUBMED 10593569
 REFERENCE 2 (bases 1 to 1656)
 AUTHORS Shi,J.D., Wang,C.Y., Marron,M.P., Ruan,Q.G., Huang,Y.Q.,
 Dettler,J.C., Davoodi-Semirani,A. and She,J.X.
 TITLE Complete genomic sequence, gene structure and localization of the
 mouse Aire gene
 JOURNAL Unpublished
 AUTHORS Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
 TITLE Direct Submission
 JOURNAL Submitted (17-FEB-1999) Pathology, University of Florida, 1600 SW
 Archer Road
 FEATURES
 source Location/Qualifiers
 1..1656
 /organism="Mus musculus"
 /strain="SJL"
 /db_xref="taxon:10090"
 /chromosome="10"
 /map="between D10Mit31 and D10Mit10"
 1..>1656
 /gene="Aire"
 1..>1656
 /gene="Aire"
 /function="transcription factor"
 /product="autoimmune regulator"
 /protein_id="AF36481.1"
 /db_xref="GI:7108573"
 /translation="MAGDGMRLRLRLHRTETAVADSAFPLLIHALADHDVPEDEK
 OETLRKEBEGQAFHALISWLTDSGAILDFWRLFRDYNLEYSRLHSILDFP
 KVDVLDNOSRRKRPPLAGPRAVAVLPPTPRKALEPRATPATLASVSPSHLK
 TPRPKPDNLESOHLPLNGIQDMAVSQRAVVASGVDPATGAVGAILIOVFPSS
 GRSKCIQVGEFYTPNKFEDPSGNLKNKRSGSSSLKPVYRAGAGVATIGRDOKXG
 OOCGVPLPSLEPQVOKNEDECAVCHHGELICDGCPRAFHLCLSPDLETFS
 GIMRSCCCQAGROOHLSDPEVSRPELREFTPLVLKNSASETRGSPRLKASDA
 AVTYVNLHPHAPALPEPSALCPILSLAGNEGRGPAISARCSVCGDGTGLRCAHCA
 AAFHMCHPPTAAARPTNLRCSCSADSPPTGTGAEVPTSGPRAPGLAKVGDSDS
 ASHPVLIHRDLESLINLEHSDFILOWAIDMSRPLAETPPFS"
 BASE COUNT 344 a 532 c 479 g 301 t
 ORIGIN
 Query Match 42.9%; Score 964.2; DB 10; Length 1656;
 Best Local Similarity 76.0%; Pred. No. 1.7e-160;
 Matches 1249; Conservative 0; Mismatches 373; Indels 21; Gaps 4;

429 TGTGACCTTCAGCCAGCCCCGGAAGGGAGAGAACCCCGGCTGTCCCAAGGCTTTGGT 488
 312 TGTGAGACTTAACACAGTCCCGGAAGAGAGAAAGACCCCTGTGTCCCAAGGCGGGGT 371
 489 ACCGCCACCAAGACTTCCCAACCAAGAGAGAGGCTTCAGAGAGGCTGAGCTGCCGGCC 548
 372 ACTGCCACCAAGAGACCCCAACCAAGAGAGAGAGGCTTCAGAGAGGCTGAGCTGCCGGCC 431
 549 AGCAGCCCTGACTTCAG 608
 432 AGCAACTCTGCGCTCAAG 491
 609 CAAAGAGCCGAG 668
 492 TAAGAGCCAGAGATGGCAACTTGGAGTACAGACACTTCTCTTGGAAACGAATTTCAGAC 551
 669 CATGTAGCTTACATGCTCAG 728
 552 CATGCGAGCTTCTGTCTCAG 611
 729 AGGGGCGTGGAG 788
 612 AGGGGCGTGGAG 671
 789 CATCCAGTTGGTGGAG 845
 672 CATTCAGTTGGTGGAG 731
 846 GAAGAACAAAGGCGCCGACAG 905
 732 GAAGAACAAAGGCGCCGACAG 791
 906 GGGGCGCTGCCCGCGTGGAG 965
 792 GGTACTAATACCTGGTAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 851
 966 TCTGGCCCTCCCGCAGTACAGCCCGCAGCTCCACCAAGAAAGAGAGAGAGAGAGAGAGAGAG 1025
 852 TCCATCCCTCCCGCAGTACAGCCCGCAGCTCCACCAAGAAAGAGAGAGAGAGAGAGAGAGAG 911
 1026 TCGGAGAGCGGGGAGCTCATCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1085
 912 CCAGCAGGAGAGTGAACCTATCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 971
 1086 CCGTCCCTCCCGCAGTACAGCCCGCAGCTCCACCAAGAAAGAGAGAGAGAGAGAGAGAGAG 1145
 972 CCGTCCCTCCCGCAGTACAGCCCGCAGCTCCACCAAGAAAGAGAGAGAGAGAGAGAGAGAG 1031
 1146 GGCACAGTCCAG 1205
 1032 GGGCAGAGTCCAAAG 1091
 1206 GGAAGACCCCGTCCCGCGGGGCTTAAGTCCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1265
 1092 AGAGACCCCGATCTCTGTGGAGTGAAGTCACTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1151
 1266 GGAACCCCTAGCGCGGATGAGACAGACTCTGTGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1325
 1152 GAGAGTCAAGGAG 1211
 1326 TGCAGCCCGCTGCGAG 1385
 1212 TGCAGCT 1265
 1386 TGAAGGTAGCAGAACTGTGCTCTGTGTGCTGTGTGCTGTGTGCTGTGTGCTGTGTGCTGTGTG 1445
 1266 TGAAGGTAGCAGAACTGTGCTCTGTGTGCTGTGTGCTGTGTGCTGTGTGCTGTGTGCTGTGTG 1325
 1446 GGTGTGCGGTACTCACTGTGCGCGGCTGTCTCACTGTGCGCGGCTGTCTCACTGTGCGCGGCTGT 1505
 1326 GGTGTGCGGTACTCACTGTGCGCGGCTGTCTCACTGTGCGCGGCTGTCTCACTGTGCGCGGCTGT 1385
 1506 CACCTCCCGGCGCGGAG 1559

QY	1146	GGCAACAGTCACGAGAGGTGACGCCCGGACAGAGAGACCCCGCCAGACACCCTT	1205
Db	1032	GGGACAGATGTCACACAGAACTGTGTCCACAGCTTGAGGTGTCCAGGACCCCGGAGCTTACCTGC	1091
QY	1206	GGAGACCCCGCTCCCGCGGGGCTTAGTGGGCGGAGAGAGAGTAAAGAGTGCACCTGG	1265
Db	1092	AGAGACCCCGCTCTGTGTGGACTGAGGTACGCTTCAGAGAAAAACAGGGGCCCTATCCAG	1151
QY	1266	GGAAACCCCTTAGCCGCGCATGGACACGACTTGTGTACAAAGACACTGCCGGCTCCGGCTTC	1325
Db	1152	GGAGCTCAAGGCCACAGCTCTGTATGCTGTGCATATATGTAACTGTGATGCTGACCCCGACCC	1211
QY	1326	TGAGACCCCGCTCCAGAGGCTGGAGTCTCTGGCCGCGGCACACCCCTACTGTGTGGTGC	1385
Db	1212	TGAGACTCTCTCT-----GTTGAGCCCTTACGACACTGTGTCCCTCTTACTGTAGTCTTGGA	1265
QY	1386	TGAGGGTCAAGACAGACACTGTGCTCTGTGGCGCTTGTGGGATGTGGCAGATGTTACGGA	1445
Db	1266	TGAGGGGCGGCGCAGGTTCAGACACCAAGCGCGGATGACAGTGTGTGTGGCATGTGGACCGA	1325
QY	1446	CTGTCTGGCGTGTATCTACTGCGCGCTGTCTTCCACTGGCGCTGCACCTTCCAGCCGG	1505
Db	1326	GGGTGTGCGGTGTGCACACTGTGCTCCCTCTCTTCCACTGGCGCTGCACCTTCCAGCGCG	1385
QY	1506	CACCTCTCGGCGCGGACGAGGACTGTGGCTGTGACATCTCTGTACAGAGAGTGT-----AC	1559
Db	1386	CGCGGCGCGCGCGCGGACCAATCTCCGCTGTCAATCTCTGTGTGCAGACTGTCACTTCCAC	1445
QY	1560	CCGACGCCCTGTGTGAGGGGTGCTGTGGCCCGCCAGCCCGCGCGCTGTGGCCCTGTGGCGCTAC	1619
Db	1446	GGCAGGCGACACCGGGGAGAGCTGTACCCACACTCTGTGGGCGCCGCGCAGCACTGGGGTTGC	1505
QY	1620	CAA-----GGATGACACAGCCAGTACAGAGCCCGCTGTCTGCACAGGATGACCTGGAGTC	1673
Db	1506	CANAGTAGGGGAGGACTGT	1565
QY	1674	CCTTCTGAGCGAGACACACCTTGGATGGCATCTGTGAGTGGGCGCATTCAGAGATGGGCGCG	1733
Db	1566	CCCTCTCATATGAGCACTCATTTGAGGGCATCTCTGAGTGGGCGCATTCAGAGATGTGCACG	1625
QY	1734	TCCGGCGCGCCCTTCCCTCCT 1756	
Db	1626	CCCGCTGCGCGAGACACACCT 1648	
RESULT 8			
LOCUS	MMU132243		
DEFINITION	Mus musculus mRNA for Aire protein.	1659 bp	linear
ACCESSION	AJ132243		
VERSION	AJ132243.1	GI:4456674	
KEYWORDS	Aire gene; Aire protein.		
SOURCE	Mus musculus.		
ORGANISM	Mus musculus.		
REFERENCE	Enkayota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1659)		
TITLE	Bleaschmidt, K., Schweiger, M., Wertz, K., Poulson, R., Christensen, H.M., Rosenthal, A., Lehrach, H. and Yaspo, M. L.		
JOURNAL	The mouse Aire gene: comparative genomic sequencing, gene organization, and expression		
MEDLINE	Genome Res. 9 (2), 158-166 (1999)		
PUBMED	99148139		
REFERENCE	10022980		
AUTHORS	2 (bases 1 to 1659)		
TITLE	Direct Submission		
JOURNAL	Submitted (14-JAN-1999) Karin B., Max Planck Institute for Molecular Genetics, Ihnestrasse 73, Berlin D-14195, Germany		
FEATURES	Location/Qualifiers		
SOURCE	1..1659		
	/organism="Mus musculus"		

[illegible]


```

Db 372 ACTGACACCCAGACCCACCCACAGAGAAAGACTGGAGGAGCTCGAGCCACCCACCC 431
QY 549 AGCAGCCCTGACTCCAGAGGGGACCCAGCCAGCTCTCAACTGAAGGCCAGCCCG 608
Db 432 AGCAACTCTGCTCAAGAGAGGTCTCAGAGCCAGGCTCCACCTGAAAGACTAATGCCC 491
QY 609 CAAGAAGCCGAGAGACGACGAGAGAGAGGCGCTTCCACTGCGGAAGGGAATTCAGAC 668
Db 492 TAAAGACGCAATGGCAACTTGGAGTACAGACCTCTCTTGGAAACGGAATTTCAGAC 551
QY 669 CATGTCAAGCTTCATCCAGAGAGGCTGTCCTCCGAGGAGTCCGAGGAGCCCG 728
Db 552 CATGCGAGCTTCTCTCCAGAGAGCTGTACCTGCTCTGAGGATGTTCCAGAACTCG 611
QY 729 AGGGCCGTGGAGGGATCTCTATCCAGAGAGTGTGAGTACAGCGGCTCCAGAAAGT 788
Db 612 AGGGCCGTGGAGGGATCTCTATCCAGAGAGTGTGAGTACAGAAAGTCCAGAAAGT 671
QY 789 CATCAGAGTGTGGAGAGTCTACACTCCAGCAAGTTCAGAACTCC--GGCAGTGG 845
Db 672 CATTCAGAGTGTGGAGAGTCTATACACCCAGAAAGTTCGAAAGCCAGTGGCAATT 731
QY 846 GAAGAAACAGGCGCCGACAGACAGTGGCGCCGAGGCTCTGGTGGAGCAAGGAGCCCA 905
Db 732 GAAGAAACAGGCGCCGAGAGTGTACAGCTTAAAGCCAGTGTCCAGCCAGAGGAGCCCA 791
QY 906 GGGCGCTGCCCCCGGTGGAGGTAGGCTGAGGCGCCAGCAGGAGCGTTCGCGCCCG 965
Db 792 GGTCTACTATACCTGTAGATGAGATGAGAAAGTGGGCGCAGAGTGTGGGGTTCCTCCCT 851
QY 966 TCTGGCCCTCCAGAGTACCCCGCTCCACCAAGAAATAGAGACGAGTGTGCGCTGTG 1025
Db 852 TCCATCCCTCCAGAGTACCCCGCTCCAGGTTAAACCAAGAAAGATGATGTGCGCTGTG 911
QY 1026 TCGGAGAGGCGGGGAGTCTATCTGTGAGCGGCTGCGGCGCTTCCAGCTGGCGCTG 1085
Db 912 CCAGAGAGGAGGTGAGTCTATCTGTGAGCGGCTGCGGCGCTTCCAGCTGGCGCTG 971
QY 1086 CCTGTCCCTGCGCTCCGAGAGATCCCAAGTGGAGCTGGAAGTCTCCAGCTGGCTGCA 1145
Db 972 CCGTGTCCCTGCGCTCCGAGAGATCCCAAGTGGAGCTGGAAGTCTCCAGCTGGCTGCA 1031
QY 1146 GGCAGACAGTCCAGAGGTGAGGCGCCGCGGAGAGAGAGTAAAGAGTCCAGCTG 1205
Db 1032 GGGCAAGTCCAGAGAGTCTGTCAGCTGAGTGTCCAGGCTGAGGAGTCCAGCTG 1091
QY 1206 GGAAGACCCGCTCCCGCGGCTTGAAGTGGGAGAGAGTAAAGAGTCCAGCTG 1265
Db 1092 AGAGACCCGAGTCTGTGGAGTGAAGTCTGCAAGAAACCAAGGAGCCATCCAG 1151
QY 1266 GGAACCCCTAGCCGAGTGAAGACGAGTCTGTCTACAGACACCTGCGGCTCCAGCT 1325
Db 1152 GGAAGCTCAAGAGCAGCTGTGATCTGTCTGACATATGTAACCTGCTGGCGCCAGCC 1211
QY 1326 TGCAGGCGCGCTGCGAGGCTGAGTCTGCGGCGCTGACCCCTACTGTGTGGGCTC 1385
Db 1212 TGCAGGCTCTCTT-----GCTGAGGCTTCAAGCAGTGTGCTACTGTGAGTGTGGGAA 1265
QY 1386 TGAGGTTGAGAGAACTGCTCTGTGCGGCTGAGGAGTGGTGAAGTGAAGTGA 1445
Db 1266 TGAGGAGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1335
QY 1446 CGTGTGCGGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1505
Db 1336 GGTGTGTGGGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1385
QY 1506 CACCTCCGAGGCGCGGAGAGGAGGCTGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1559
Db 1386 GCGCCCGGAGGAGGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1445
QY 1560 CCGAGCCCTGTGAGAGGAGTGTGGCGCCGAGCCCGCGGCGCTGGCGCTGCG 1619

```

```

Db 1446 GCCAGGACACCGGCGGAAGCTGTACCCACTCTGGGCGCCGTCAGCAGCCTGGCTGC 1505
QY 1620 CAA-----GGATGACATCGCATCAAGAGCCCGGCTGTGACAGGATGACCTGATGC 1673
Db 1506 CAAGGTAGGGGACACCTGTCTGATCAAGACCTGTGTCTTACATGAGGAGCAGCTGAGATG 1565
QY 1674 CATTCTGAGCGAGACACCTTGTGATGATCTCAGTGGGCGCATTCAGACATGAGCCG 1733
Db 1566 CTTCTCAATGAGAGACTCATTTGACGGCATCTCAGTGGGCGCATTCAGACATGATCAGC 1625
QY 1734 TCCGGGCGGCGCCCTTCCCTCTCT 1756
Db 1626 CCGCGTGGCGGAGACACCACTT 1648

RESULT 10
AF079536
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus.
Mus musculus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1936)
Wang, C. Y., Shi, J. D., Davoodi-Semiromi, A. and She, J. X.
Cloning of Aire, the mouse homologue of the autoimmune regulator
(AIRE) gene responsible for autoimmune polyglandular syndrome type
1 (ASPI)
JOURNAL
Genomics 55 (3), 322-326 (1999)
MEDLINE
99168902
PUBMED
10049587
REFERENCE
2 (bases 1 to 1936)
Wang, C. Y., Shi, J. D., Davoodi-Semiromi, A. and She, J. X.
Direct Submission
Submitted (21-JUL-1998) Pathology, Immunology and Laboratory
Medicine, University of Florida, 1600 SW Archer Road, Gainesville,
FL 32610, USA
FEATURES
source
location/Qualifiers
1..1936
/organism="Mus musculus"
/strain="B6"
/db_xref="taxon:10090"
/tissue_type="thymus"
1..1936
/gene="Aire"
58..1716
/gene="Aire"
/function="transcription factor"
/product="autoimmune regulator"
/protein_id="A02044.1"
/db_xref="gi:4426599"
/translation="MAGSDGMLRLLRLRHTREIAVAIDSAFLLHALADHDVPEDEK
OETLRKEKREGCPQAFHALLSMILTRDSGALIDFMRILFKOVNLERYSRYSITLGGF
KDVLDNRKGRKPLAGKAVLPRPTTKRALEPRATPATLASSVSSPGSHLK
TKPRKPKDNGLESQHLPLNGKIQTMAASVQAVVAVASDVPRGAVGILIQVEES
GRSKKCIQVGGERTYTPNKFEDPSGMLKKRASGSLKPYRAKGAQVITPGRDEKVG
QOCCVPPLPLSEPOYNOKNEDBCAVCHDGELICDGCPRAFILACLSPIQEIPI
QIMRSCCLQGRVQONLSQPSRPELPATPIIVGARSASEKTRGPRSELKASSDA
AVTYVNLAPHPAPLLEPSALCPILLSNGNRRGPRGAPASCSNGDTEYLKCHCA
AAFWHRHPTAARPGTNLCKSCSADSTPTPTGPAVPTSGPRAPGLAKVGDDDS
ASHPDVHRRDDELNLNHSFDGILQWAIQSMRPLAETPPSS"
BASE COUNT
408 a 609 c 568 g 351 t
ORIGIN
Query Match 42.9%; Score 964.2; DB 10; Length 1936;
Best Local Similarity 76.0%; Pred. No. 1.6e-160;
Matches 1249; Conservative 0; Mismatches 373; Indels 21; Gaps 4;
QY 129 GGACGGGCGCTTACGCCGCTTGTAGGCTGACACCGACGAGATCGCGGTGCGGTGA 188

```

Db 69 GATGAAATGCTACGCGCTGCTGAGGTGACCGACCGAGATGCGGTGCGCATAGA 128
 QY 189 CAGCCCTTCCCACTGCTGACAGCGCTGCTACCAAGAGAGTGTCCCGAGCAAGTT 248
 Db 129 CAGTCCCTTCCCGCTGCTGATGCTCTAGCCACCAAGAGAGTGTCTCTGAGAGCAAGTT 188
 QY 249 TCAGAGAGCGCTTCACTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 308
 Db 189 CCAGAGAGAGCGCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 248
 QY 309 GTCCGTGCTGCTGAG 368
 Db 249 GTCCGTGCTGCTGAG 308
 QY 369 GAGCTACAACTGAG 428
 Db 309 GAGCTACAACTGAG 368
 QY 429 TGTGAGACTTACAG 488
 Db 369 TGTGAGACTTACAG 428
 QY 489 ACCGCAACCCAG 548
 Db 429 ACTGCAACCCAG 488
 QY 549 AGCAGCCCTGAG 608
 Db 489 AGCAACTCTGAG 548
 QY 609 CAAAG 668
 Db 549 TAAAG 608
 QY 669 CATGTCACTTCACTGAG 728
 Db 609 CATGTCACTTCACTGAG 668
 QY 729 AGGGGCGGTGAG 788
 Db 669 AGGGGCGGTGAG 728
 QY 789 CATCAAGTGTGAG 848
 Db 729 CATCAAGTGTGAG 788
 QY 846 GAAG 908
 Db 789 GAAG 848
 QY 906 GGGGCGGTGAG 968
 Db 849 GGGGCGGTGAG 908
 QY 966 TGTGAG 1028
 Db 909 TGTGAG 968
 QY 1026 TGTGAG 1088
 Db 969 TGTGAG 1028
 QY 1086 CCGTGTGAG 1148
 Db 1029 CCGTGTGAG 1088
 QY 1146 GCGAAG 1208
 Db 1089 GCGAAG 1148
 QY 1206 GCGAAG 1268
 Db 1206 GCGAAG 1206

Db 1149 AGAGACCCCGATTCCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1208
 QY 1266 GGAACCCCTTACCGGAG 1325
 Db 1209 GGAACCCCTTACCGGAG 1268
 QY 1326 TGCAG 1385
 Db 1269 TGCAG 1322
 QY 1386 TGCAG 1445
 Db 1323 TGCAG 1382
 QY 1446 TGCAG 1505
 Db 1383 TGCAG 1442
 QY 1506 TGCAG 1563
 Db 1443 TGCAG 1502
 QY 1560 TGCAG 1619
 Db 1503 TGCAG 1562
 QY 1620 TGCAG 1673
 Db 1563 TGCAG 1622
 QY 1674 TGCAG 1733
 Db 1623 TGCAG 1672
 QY 1734 TGCAG 1795
 Db 1683 TGCAG 1705
 RESULT 11
 AF128115
 LOCUS
 DEFINITION
 Mus musculus autoimmunity regulator (Aire) mRNA, alternatively spliced product 1b, complete cds.
 AF128115
 AF128115.1 GI:7108531
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1933)
 Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
 Expression and alternative splicing of the mouse autoimmunity regulator gene (Aire)
 Unpublished
 2 (bases 1 to 1933)
 Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
 Direct Submission
 Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW Archer Road, Gainesville, FL 32610, USA
 Location/Qualifiers
 1..1933
 /organism="Mus musculus"
 /strain="B6"
 /db_xref="taxon:10090"
 /chromosome="10"
 /map="between D10Mit31 and D10Mit10"
 1..1933
 /gene="Aire"
 58..1713
 /gene="Aire"
 /function="transcription factor"

[illegible]

QY	1206	TCGGGAGGGGGGAGCTCATCTGCTGTAGCGGTGCCCTTCGGGCTTCCACCTGGGCTG	1085
QY	966	TCCTGGCCCTCCCAAGTACGCCCCAGCTCCACACAGAAATGAGGACGATGTGCCGTGTG	1025
Db	849	GCTCACTATACCTGTGTAGATGTAGCACAAAGTGGGCGACGATGTGGGCTTCTCCCTT	908
QY	1026	TCGGGAGGGGGGAGCTCATCTGCTGTAGCGGTGCCCTTCGGGCTTCCACCTGGGCTG	1085
Db	909	TCCATCCCTCCCAAGTGGGCCCAAGGTTAACCAAGAC---GAGGATGAGTGTGCCGTGTG	965
QY	1086	CCTTTCCTCCCTCCGCTTCGGGGGATCCCCAGGGGACCTTGGAGGTGTCACGCTGCCGCA	1145
Db	1026	CTTCTCCCACTTGTGAGAGGATCCCCAGTGGCTTGGAGATGCTCTGTGCTCCCA	1085
QY	1146	GGCAACAGTCCAGAGAGGTGACGCCCCCGGACAGAGAGACCCCGCCAGAGACCCGT	1205
Db	1086	GGGACAGATCCAAAGAAACCTGTGCCAGCCGTGAGAGTCCAGAGGCCCGGACGCTACCTGC	1145
QY	1206	GGAGACCCCTCCCTCCCGGGGCTTTAGGTGGCGGGAGAGAGAGTAAAGATCCACCTGG	1265
Db	1146	AGAGACCCCATCTCTCTGGAGCTGAGAGTCAAGCTTCAAGAAAAACAGGGGCCATTCAG	1205
QY	1266	GGAAACCCCTCCCGGACCTGGACAGACTCTTGTCTAAACACCTCCGCGCTCCGCTTC	1325
Db	1206	GGAGCTCAAAAGCCAGCTCTATGCTGCTGTACATATAGTAAACCTGCTGGCCCGCACCC	1265
QY	1326	TGCAGCCCGCTCCAGAGGCTGGACTTCCTCGGCCCTGCACACCCCTACTGTGTGGGTCC	1385
Db	1266	TGCAGCTCTCTCTCTCTGGAGCTTCAAGCAGTGTGCTCTACTGTGTGGGAA	1319
QY	1386	TGAGGCTCAGCACAACCTGGCTCTCTGTGCGGCTTGGGGGTGTGGAGATGTACGA	1445
Db	1320	TGAGGGGGCGGCGAGTTCACAGCAACAGCGCGCATCAGTGTGTGGCATGGACCGA	1379
QY	1446	CGTGTCCGCTTACTACTGCGCGCTGCTTCACATGGCGCTGGACACTTCCACACCG	1505
Db	1380	GGTGTGTGGCTGTGCACTGTGCTGCTTCCATGGCGCTGCACACTTCCGACGGC	1439
QY	1506	CACCTCCGCGCGGAGGAGGCTGCGCTGACAGATCTCTCAGAGACGTG-----AC	1559
Db	1440	CGCGCGCGCGGCGGAGCAACATCTCGGTGCAAAATCTCTGTGCAAGACTCAGCTCCAC	1499
QY	1560	CCGAGCCCTGTGTGAGGGGTGCTGGCCCCCAGCCCCCGCGCTGGGCTGTC	1619
Db	1500	GGCAGGCAACCGGGGAACTGTACCCACCTCTGGGGCCCGCTCCAGCACTGGGCTTGC	1559
QY	1620	CAA-----GATGTACATCCGACAGACAGGCGCGCTGTGCAAGGAGTACCTGGAGTC	1673
Db	1560	CAAGGTAGGGGAGCACTGTGCTAGTCACACCTGTCTTACATTAAGGAGACACTTGGAGTC	1619
QY	1674	CCTTGTGAGCAGACACACCTTCGATGGCATTCCTGTGACGTGGGCAATCCAGAGCATGGCCG	1733
Db	1620	CTTCTTCATGTAGCACTATTTTGAAGGCACTCTGCAGTGGGCAATCCAGAGCATGTACAG	1679
QY	1734	TCCGGCGGCGCCCTTCCCTCTCT	1756
Db	1680	CCCGCTGGCGGAGACCAACCTT	1702

RESULT	12
AFI28116	
LOCUS	
DEFINITION	1924 bp mRNA linear ROD 29-FEB-2000
ACCESSION	Mus musculus autoimmune regulator (Aire) mRNA, alternatively spliced product 1c, complete cds.
VERSION	AFI28116
KEYWORDS	AFI28116.1 GI:7108533
SOURCE	.
ORGANISM	Mus musculus. Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
1 (bases 1 to 1924)
AUTHORS
Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.
TITLE
Expression and alternative splicing of the mouse autoimmune regulator gene (Aire)
JOURNAL
Unpublished
2 (bases 1 to 1924)
AUTHORS
Ruan, Q.G., Wang, C.Y., Shi, J.D. and She, J.X.
TITLE
Direct Submission
Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW
Archer Road, Gainesville, FL 32610, USA
FEATURES
source
1. 1924
/organism="Mus musculus"
/strain="B6"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1. 1924
/gene="Aire"
58. 1704
/gene="Aire"
/function="transcription factor"
/note="alternatively spliced product 1c"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AF36461.1"
/db_xref="GI:7108534"
/translation="MAGDGLRLRLHRTAVLIDSAFPLIALADHVDK
OETLRKEGCPDAPFALISMLTIRDSGLIDPWRLFDYNERLSILIDGP
KDYLDNOSRGRPLAGPKAVLPFRPTKALIEPRATPATLKSVSPSHLK
TRPKRFDGLSOLPLNGIDTMAVSQRAVAVAGDVGTGAEGILIQVFES
GRSKICIQVEGEYTPNKEFDPGSKNKNRSGSLPVAFAKAGDEKVGQCG
VPLPLPSEPPOYNKDECAVCHDGEICDCEPRAFHACTLSPLOIPGLMR
CSGCLGVRVOONLSQEVSRPELPETPLIVGGRSSEKTRGSRLEKASDAVY
VNLADPAPALIEPALICPLISAGNEGRGAPASCSUGDBTEVLRCHKAARH
WRCHPEAARPTNLRCKSCSDSTPTPTPGAVTSGFRPAPGLAKVEDSASHD
PVLHRDLSLNEHSPDGLIOWAIOISMPLAETPPSS"

BASE COUNT 405 a 605 c 567 g 347 t
ORIGIN

Query Match 41.9%; Score 939.8; DB 10; Length 1924;
Best Local Similarity 75.7%; Pred. No. 3.3e-156;
Matches 1243; Conservative 0; Mismatches 367; Indels 33; Gaps 5;

QY 129 GACGCGGCGCTACGCGGCTTGTGAGCTGCACGACGAGATCGCGGTGCGCTGA 188
DB 69 GATGGAATGCTACGCGGTGTGCTGAGGCTGCACCGACGACGATCGCGGTGCTATGA 128
QY 189 CAGCGCTTCCACTGCTGCACGCGCTGCTGACACGACGTGCTCCCGAGGACAGTT 248
DB 129 CAGTGCCTTCCGCTGCTGATGCTTACGACGACGACGTGCTCCCTGAGGACAGTT 188
QY 249 TCAGGAGCGCTTCATGTGAAGAAAAGAGGCGTGCCTCCAGGCTTCCAGGCTTCT 308
DB 189 CAGGAGAGCGCTCGCTGTGAAGGAGAAAGGCTGCTCCCGAGGCTTCCAGGCTTCT 248
QY 309 GTCTGCTGCTGACCGACGACCTCAACGACCATCTGACTTGTGAGAGGCTGTTCAA 368
DB 249 GTCTGCTGCTGACCGCGGACGACGACATCTGATTTCTGAGAGATTTCTTTTAA 308
QY 369 GACTTCAACCTTGGAGCGCTATGCGCGGCTGCAGCGCATCTTGACAGCTTCCCAAGA 428
DB 309 GGACTCAACATCTGGAGCGGTACAGCGCTGATAGCATCTTGAGCGGCTTCCCAAGA 368
QY 429 TGTGACCTCAGCGACCGCGGAGGAGAGCGCGCGCTCCCGCAAGGCTTTGGT 488
DB 369 TGTGACCTCAACCACTCCCGGAGAGGAGAAAGCGCTTGTGCTGCTCAAGCGCGGCT 428
QY 489 ACCGCGACCCAGACTCCGACCAAGAGAGAGGCTCAGAGAGGCTGAGCTGCGCGGCGC 548
DB 429 ACTGCGACCCAGACCCCGACCAAGAGAGAGGCTGAGAGGCTGAGCGCACCCACACC 488

QY 549 AGCAGCCCTGACTCCCAAGGGGACACCGCCAGCGCTCTCACTGAGAGCCAGCCCGC 608
DB 489 AGCAACTCTGCGCTCAAGAGCGCTTCCAGCGCCAGCGCTCCACCTGAAAGATTAAGCCCGC 548
QY 609 CAGAAAGCCGAGAGACAGCCGACAGCAGCGCTTCCACTCGGGAAGGAGATTACAG 668
DB 549 TAAGAGCCGAGATGCGCAACTTGGAGTCACAGCACCTTCTCTTGGAAACGAAATTACAG 608
QY 669 CATGTCAGCTTCACTGACAGAGAGCTGCGCATCTCCCGGGAGACCGCGGAGCCCG 728
DB 609 CATGTCAGCTTCTGTCTCAGAGAGCTGTACCGCTGTGGGATGTTCCAGAAACCGG 668
QY 729 AGGCGCGTGGAGAGGATCCCTATCCAGAGGTTTGTAGTCAGCGGCTCCAGAAAGTG 788
DB 669 AGGCGCGTGGAGAGGATCCCTATCCAGAGGTTTGTAGTCAGAGATTCAGAAAGTG 728
QY 789 CATTCAGGTTGGTGGGAGTTTCACTCCAGCAAGTTTCGAGATTC--GGCAGTG 845
DB 729 CATTCAGGTTGGGAGGATTTATACACCAACCAAGTTTCGAGATTCGAGCAATT 788
QY 846 GAAGAACAAAGGCGCGGACAGCAGAGCTGCGCAACCTCTGTTGAGCCAAAGAGCCCA 905
DB 789 GAAGAACAAAGGCGCGGAGTGTAGCAGCTTAAGCCAGTGTGCGCAAGAGGAGCCCA 848
QY 906 GGGCGCTGCGCCCGGTGGAGAGTGAAGCTTAAGCTGGGACAGAGGAGCTTCCGCGCC 965
DB 849 G-----GTTAGATATGACGAGAAAGTGGCGCAGCAGTGTGCTTCCCT 896
QY 966 TCTGCGCTTCCCGACGATGACCCCGACGCTCCACCAAGATGAGAGAGTGTGCGGTG 1025
DB 897 TCCATCTCTCCCGACGATGAGCCCGAGGTTAACCAAGAAAGAGATGATGTGCGGTG 956
QY 1026 TCGGGAGGGGGGAGCTATCTGCTGTGACAGGCTGCGGCTTCCAGCTGCGCTG 1085
DB 957 CCAGCAGGAGGAGTGTGATCTATGTTGTGAGCGGTGTCGCGGCTTCCAGCTGCGCTG 1016
QY 1086 CCTGTCCCTCCGCTCCGAGAGATCCCACTGGAGCTGAGAGTGTCCAGCTGCTGCA 1145
DB 1017 CCGTCCCGACCTTCGAGAGATCCCACTGGAGTGTGAGATGCTCTGCTCTCCA 1076
QY 1146 GGCACAGTCCAGAGAGTGCAGCCCGGAGAGAGAGAGAGAGAGTGTGAGAGTGTG 1205
DB 1077 GGGCAGAGTCCAAACAAACACTGCTCCAGCTGAGAGTGTGAGAGGCGCGGAGTCACTGC 1136
QY 1206 GAGAGACCGGCTCCCGCGGCTTGTAGTGGGAGAGAGAGTGTGAGAGTGTG 1265
DB 1137 AGAGACCGGATCTCTGTGGAGTGAAGTCAAGTTCAAGAAACCAAGGCGCATCAG 1196
QY 1266 GGAACCCCTAGCGGATGAGACAGACTTGTCTAACAGCAGCTGCGGCTCGGCTTC 1325
DB 1197 GAGAGTCAAGAGCAGCTGTGATGCTGTGATGATGATGATGATGATGATGATGATGATG 1256
QY 1326 TCCAGCGCGCTGCGAGGCTGACCTCTGCGGCTGACACCCCTACTGTGTGGTTC 1385
DB 1257 TCCAGCTCTCT-----CTGAGGCTTTCAGACACTGTGCTCTACTGATGCTGGGAA 1310
QY 1386 TTAGGGTCAAGAGAACTGCTGCTGTGCGGCTTGGGAGTGTGAGAGATGTTACGA 1445
DB 1311 TTAGGGGCGGCGAGGTCAGCAACCAAGCGCGCATAGTGTGTGGCGATGGACCGA 1370
QY 1446 GGTGCTGCGTGTACTACATGCGCGCTGCTCACTGAGCGGTGCGACATTCAGCGCG 1505
DB 1371 GGTGTTGGGTTGTGACACTGTGCGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1430
QY 1506 CACTTCCCGCGCGGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1559
DB 1431 CGCGCGCGCGCGGAGCAACATCTCGCTGCAAAATCTGCTGCTGCTGCTGCTGCTGCTG 1490
QY 1560 CCGACCTCTGAGAGGAGGCTGTGCGGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1619
DB 1491 GCGAGGCAACCGGCGGAGAGTGTACCTCTGTGGCGCGCGCGCGCGCGCGCGCGCGCG 1550
QY 1620 CAA-----GGATGACACTGTCCAGTCAAGGCCCGGCTGTGACAGGAGATGACTGAGTC 1673


```

|||||
Db 1214 GGGCATGCAAGCTGTGATGAGAGAGTGTGAGAGAGACACCTCTCTCTCATCTGGAA 1273
QY 1831 GCGGCGCGGCTGGGATCAAGAGGAGGAGCGCCACCTCTGTGTGCTGGCTGTAA 1890
Db 1274 GCGGCGCGGCTGGGATCAAGAGGAGGAGCGCCACCTCTGTGTGCTGGCTGTAA 1333
QY 1891 CAGCTGTGTGTCTGGGAGACACGACATGCTGTGAAATTAACCTGGCCGAC 1950
Db 1334 CAGCTGTGTGTCTGGGAGACACGACATGCTGTGAAATTAACCTGGCCGAC 1393
QY 1951 TTCTACTGTGAAGTCCCGGAGGAGCTCTGCTGTGAGCTACTAAATATA 2010
Db 1394 TTCTACTGTGAAGTCCCGGAGGAGCTCTGCTGTGAGCTACTAAATATA 1453
QY 2011 AAATTAGCTG 2020
Db 1454 AAATTAGCTG 1463

RESULT 15
AF128122 1637 bp mRNA linear ROD 29-FEB-2000
LOCUS AF128122
DEFINITION Mus musculus autoimmune regulator (Aire) mRNA, alternatively
ACCESSION AF128122
VERSION AF128122.1 GI:7108545
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1637)
AUTHORS Ruan,Q.C., Wang,C.Y., Shi,J.D. and She,J.X.
TITLE Expression and alternative splicing of the mouse autoimmune
regulator gene (Aire)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1637)
AUTHORS Ruan,Q.G., Wang,C.Y., Shi,J.D. and She,J.X.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1999) Pathology, University of Florida, 1600 SW
Archer Road, Gainesville, FL 32610, USA
FEATURES
source location/Qualifiers
1..1637
/organism="Mus musculus"
/strain="B6"
/db_xref="taxon:10090"
/chromosome="10"
/map="between D10Mit31 and D10Mit10"
1..1637
/gene="Aire"
/cds 58..1287
/function="transcription factor"
/note="alternatively spliced product 3a"
/codon_start=1
/product="autoimmune regulator"
/protein_id="AAF36467.1"
/db_xref="GI:7108546"
/translat="MAGGDMRLRLRLHRTETAVAIIDSAFPLMLADHDVPEDEK
OETLRLEKEGCPDAFHALLSLTIDSGAILDFMRLIFDYULERYSLRSLIDGFP
KDVLDNOSRKGRKPLAGPAVALPRLPRKRLLEPRATPRPATLASKSVSPGSHIK
TKPKKPDGNSLESHLPLGNGIOTMAASYORAVTVASGVPTGRGAVEGLITQOVES
GRSKCIQVGEFFTPNKRKEDPSGINKLNKARSGLKPVVRAAGAVTTPGRDQKVG
QDCGVPLPSLSEPOVNOKNDECAVCHDGGELICDDCPRAFHLACLSPLQETPS
GLMRSCLQAGRVQNLISQPEVSRPELPATPDSPLDILICRLDLSHNRGSCTH
LWAPSTWACQGRRLC"
BASE COUNT 359 a 511 c 472 g 295 t
ORIGIN
Query Match 32.9%; Score 737.6; DB 10; Length 1637;
Best Local Similarity 79.9%; Pred. No. 1.5e-120;
Matches 982; Conservative 0; Mismatches 219; Indels 3; Gaps 1;

```

```

QY 129 GAGCGCGCCCTACCGCGCTTCTGAGGCTGACCGCAGAGATCGCGGTGCGCTGCA 188
Db 69 GGATGGAAATGCTACGCGCTTCTGCTGAGGCTGACCGCAGAGATCGCGGTGCGCTGCA 128
QY 189 CAGCGCTTCCACTGCTGTCAGCGCGGTGACACAGAGAGTGTCCCGGAGCAAGTT 248
Db 129 CAGTGGCTTTCCTGCTGCTGATGCTTACCGGACACAGAGTGTCCCTGAGCAAGTT 188
QY 249 TCAGAGAGCGCTTCATCTGAAGAAAAGAGGCGTCCCGCAGCGCTTCCAGCGCTTCT 308
Db 189 CCAGAGAGCGCTCGTCTGAAAGAGAAAGAGGCTGCCCGCAGCGCTTCCAGCGCTGCT 248
QY 309 GTCTGCTGCTGCTACCGAGACTCCACAGCATCTTGTGACTTCTGAGAGGCTGCTTCAA 368
Db 249 GTCTGCTGCTGCTACCGAGCGCTGAGGCGGCTGAGATTCCTGAGAGATTCCTTAA 308
QY 369 GGACTCAACCTGAGAGCGCTATGCGCGGCTGACGCGCATCTGAGAGATTCCTCCCAAGA 428
Db 309 GGACTCAACCTGAGAGCGCTATGAGCGGCTGACGCGCATCTGAGAGATTCCTCCCAAGA 368
QY 429 TGTGACCTTCAGCAGCCCGGAGAGGAGAGAGCCCGCGTCCCAAGCGTTTGT 488
Db 369 TGTGACCTTCAGCAGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428
QY 489 ACCGCAACCCAGACTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 548
Db 429 ACTGCCACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
QY 549 AGAGAGCGCTGACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 608
Db 489 AGAGAGCGCTGACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 548
QY 609 CAAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 668
Db 549 TAAAGAGCGAGATGGCACTTGGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 608
QY 669 CATGTACCTTCAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 728
Db 609 CATGTACCTTCAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 668
QY 729 AGGAGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 788
Db 669 AGGAGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 728
QY 789 CATCAGGTTGTGGGAGATTCACACTCCACAGAGTTGAGAGAGAGAGAGAGAGAGAG 845
Db 729 CATCAGGTTGTGGGAGATTCACACTCCACAGAGTTGAGAGAGAGAGAGAGAGAGAGAG 788
QY 846 GAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 905
Db 789 GAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 846
QY 906 GGGCGCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 965
Db 849 GGTCACTATACCTGTGAGAGATGAGCAGAAAGTGGGCGCAGAGTGGGGTTCTCCCT 908
QY 966 TCTGCGCTTCCCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1025
Db 909 TCCATCCCTCCCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 968
QY 1026 TCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1085
Db 969 CCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1028
QY 1086 CCGTGGCGCTCGCGCGGAGAGATCCCGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1145
Db 1029 CCGTGGCGCTCGCGCGGAGAGATCCCGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1088
QY 1146 GGCAGAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1205
Db 1089 GGCAGAGCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1148

```

Oy 1206 GGAGACCCCGCTCCCCGGGGCT 1229
 ||||||| |
Db 1149 AGAGACCCCGGACCAATCTCGCT 1172

Search completed: February 14, 2003, 15:22:42
Job time : 5667.5 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 14, 2003, 11:37:20 ; Search time 444.856 Seconds

(without alignments)
11364.871 Million cell updates/sec

Title: US-09-509-595B-1

Perfect score: 2245

Sequence: 1 cgggcgcacagcggcgcg.....ggcattcattcattgtaa 2245

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database :
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2245	100.0	2245	AAK58605	Human autoimmune p
2	2018.4	89.9	2036	AAK26936	CDNA encoding a hu
3	1025.2	45.7	1545	AAK26937	CDNA encoding a hu
4	964.2	42.9	1656	AAK58606	Mouse autoimmune p
5	851.2	37.9	1463	AAK26938	CDNA encoding a hu
6	167.8	7.5	9876	ABAL32093	Human immune syst
7	155.2	6.9	25012	ABAL32093	Human immune syst
8	153.8	6.9	5540	AAK68127	Human immune/haema
9	152	6.8	2648	AAK77183	Human immune/haema

10	152	6.8	140036	24	AAK58600	Human genomic DNA
11	151.6	6.8	26427	22	ABA20762	Human nervous syst
12	151.6	6.8	26427	22	ABA20762	Human nervous syst
13	151.6	6.8	169739	24	ABQ88186	Human osteoblast d
14	151.2	6.7	65834	22	AAK86282	Human immune/haema
15	150.8	6.7	462	22	ABL81253	Human ovarian can
16	150.6	6.7	160552	22	AAD02697	Human glycosyl sul
17	150.4	6.7	12493	22	AAK26741	Human genomic DNA
18	150.4	6.7	39328	22	ABL91800	Human lipase endo
19	150.2	6.7	555	22	AAK25120	Human breast cance
20	150.2	6.7	573	22	AAK16277	Human breast cance
21	150.2	6.7	2733	24	AAK17705	Human CDNA encodin
22	150	6.7	24934	22	AAK71542	Human immune/haema
23	150	6.7	24934	22	AAK78433	Human immune/haema
24	150	6.7	162450	21	AAK69667	Human immune/haema
25	149.6	6.7	240825	22	AAK24497	Human PG-3 gene.
26	149.6	6.7	319608	21	AAK51601	Human chromosome 1
27	149.6	6.7	319608	22	AAK509301	Human chromosome 1
28	149.4	6.7	1186	22	AAK03794	Human reproductive
29	149.4	6.7	1186	22	AAK62608	Human breast or ov
30	149.4	6.7	1186	23	ABK72133	Human ovarian anti
31	149.4	6.7	1186	24	ABK91725	Human chromosome-1
32	149.4	6.7	3071	17	AAK11778	Human prostate exp
33	149.4	6.7	3470	23	ABK25066	Human nervous syst
34	149.4	6.7	5797	22	AAK15723	Human reproductive
35	149.4	6.7	5797	22	AAK103405	Human prostate exp
36	149.4	6.7	15555	22	AAK73539	Human immune/haema
37	149.4	6.7	21278	21	AAK55965	Human immune/haema
38	149.4	6.7	145831	24	ABL62309	Human G713 3'-end
39	149.4	6.7	145831	24	ABL62309	Human immune/haema
40	149.4	6.7	145831	24	ABL62309	Human immune/haema
41	149.4	6.7	145831	24	ABL62309	Human immune/haema
42	149.2	6.6	28180	22	AAK68939	Human prostate can
43	149.2	6.6	44196	22	AAK77438	Human immune/haema
44	148.8	6.6	44196	22	AAK77438	Human immune/haema
45	147.6	6.6	13585	17	AAK11549	Human prostate exp
			445	23	ABV59216	Human prostate exp

ALIGNMENTS

RESULT 1	AAK58605	standard; CDNA; 2245 BP.
ID	AAK58605	
XX	XX	
AC	AAK58605:	
XX	XX	
XX	16-AUG-1999 (first entry)	
XX	Human autoimmune polyglandular disease type 1 (APCD1) cDNA.	
DE	Human autoimmune polyglandular disease type 1; APCD1; AIRG;	
XX	human; transcription factor; autoimmune disease; diagnosis;	
KW	Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;	
KW	APCED1; autoimmune polyglandular disease type 1; APCD1; AIRG;	
KW	human; transcription factor; autoimmune disease; diagnosis;	
KW	gene therapy; ss.	
XX	XX	
XX	XX	
OS	Homo sapiens.	
XX	XX	
XX	XX	
FT	Key	Location/Qualifiers
FT	CDS	121..1758
FT	FT	/tag= a
FT	FT	/product= "human.APCD1"
FT	FT	replace(1085,CCTG)
FT	FT	/tag= b
FT	FT	/note= "4-nucleotide insertion, results in
FT	FT	frameshift and truncated 371 aa protein"
FT	FT	replace(1090,CCTG)
FT	FT	/tag= c
FT	FT	/note= "4-nucleotide insertion, results in
FT	FT	frameshift and truncated protein"
FT	FT	replace(1284,A)
FT	FT	/tag= d
FT	FT	/note= "1-nucleotide insertion, results in

FT	frameshift and truncated 422 aa protein"
FT	replace(1365,C)
FT	/*tag- e
FT	/*note- "1-nucleotide insertion"
FT	replace(1085..1097)
FT	/*tag- f
FT	/*note- "13-nucleotide deletion, results in frameshift and truncated 372 aa protein"
FT	replace(1051)
FT	/*tag- g
FT	/*note- "1-nucleotide deletion"
FT	replace(1309)
FT	/*tag- h
FT	/*note- "1-nucleotide deletion"
FT	replace(1313)
FT	/*tag- i
FT	/*note- "1-nucleotide deletion, results in frameshift and truncated 478 aa protein"
FT	replace(889,T)
FT	/*tag- j
FT	/*note- "alters Arg codon to STOP, and truncated 256 aa protein"
FT	replace(358,T)
FT	/*tag- k
FT	replace(374,G)
FT	/*tag- l
FT	replace(1052,A)
FT	/*tag- m
FT	replace(1094,A)
FT	/*tag- n
FT	replace(708,T)
FT	/*tag- o
FT	/*note- "silent polymorphism"
FT	replace(801,T)
FT	/*tag- p
FT	/*note- "silent polymorphism"
FT	replace(1317,T)
FT	/*tag- q
FT	/*note- "silent polymorphism"
FT	replace(1698,T)
FT	/*tag- r
FT	/*note- "silent polymorphism"
FT	1995
FT	/*tag- s
FT	/*note- "Alu sequence overlapping with PKFL promoter starts at this position"
FT	2191..2195
FT	/*tag- t
XX	
XX	
XX	W09918197-A2.
XX	
XX	15-APR-1999.
XX	
XX	02-OCT-1998;
XX	98WO-EP06294.
XX	
XX	12-NOV-1997;
XX	97EP-0119810.
XX	02-OCT-1997;
XX	97EP-0111514.
XX	08-OCT-1997;
XX	97EP-0111398.
XX	
PA	(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA	(NMPU) NAT PUBLIC HEALTH INST.
XX	
PI	Aaltonen J, Bjorres P, Horelli-Kuitunen N, Lehrach H;
XX	Palotie A, Peltonen L, Perhentupa J, Yaspo M;
XX	
DR	WPI: 1999-287735/24.
XX	P-PSDB; AAY06073.
XX	
PT	New polypeptide which co-segregates in mutated form - with
PT	autoimmune polyendocrinopathy candidiasis ectodermal dystrophy
XX	
PS	Claim 1; Fig 2A; 77pp; English.
XX	

This is the nucleotide sequence of human autoimmune polyglanular disease type I (APGDI or AIRY) cDNA. It is a composite of 2 clones, which were obtained from a human adult thymus cDNA library by PCR (see AAY56609-10), and of a 3' UTR extension PCR product. The encoded protein (see AAY06073) is a transcription factor or transcription-associated factor that may associate with vimentin fibres, perhaps as part of a docking mechanism regulating nuclear translocation. Aggregates of the mutated protein may prevent formation of vimentin intermediate filaments. Mutated APGDI polypeptides co-segregate with autoimmune polyendocrinopathy candidiasis ectodermal dystrophy (APECED). The invention provides vectors and host cells for preparation of APGDI polypeptides. APECED-associated mutations found in the APGDI gene include: an insertion of 4 nucleotides (CCCTG), normally found at position 1086-1089, at nucleotide position 1085 or 1090; an insertion of an adenosine at position 1244; an insertion of a cytosine at position 1355; a deletion of nucleotides 1085-1097; a deletion of the thymidine at position 1051; a deletion of the cytosine at position 1309 or 1313; a cytosine to thymidine exchange at position 889; a guanosine to thymidine exchange at position 358; an adenosine to guanosine exchange at position 374; a guanosine to adenosine exchange at position 1052; and a cytosine to adenosine exchange to position 1094. A claimed method for testing for carriership of APECED or for the disease state involves testing for a mutation in the APGDI gene or for a mutated form of the APGDI polypeptide. APGDI nucleic acid can be used in gene therapy.

Sequence 2245 BP; 419 A; 765 C; 697 G; 364 T; 0 other;

Query Match	100.0%	Score 2245;	DB 20;	Length 2245;
-------------	--------	-------------	--------	--------------

Matches 2245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CGGCGCACAGCCGGCGCGGAGGCCACACAGCCCCGCCGGAGACCCGAGCGCAAGCGAGG	60
	1	CGGGCGACAGCCCGCGCGGAGGCCACAGCCCCGCCGGAGACCCGAGCGCAAGCGAGG	60
QY	61	GCTCCAGTGTCCCGGGAGACCAACCGGTCGGCCCAACCCCGGGTCCCGCGCCACCCC	120
Db	61	GCTCCAGTGTCCCGGGAGACCAACCGGTCGGCCCAACCCCGGGTCCCGCGCCACCCC	120
QY	121	ATGGCGACGAGACGGGCGCTACGCGCGCTTCTGAGGCTGCAGCGCAAGGATCGGCGTG	180
Db	121	ATGGCGACGAGACGGGCGCTACGCGCGCTTCTGAGGCTGCAGCGCAAGGATCGGCGTG	180
QY	181	GCGCTGGACAGCGCCTTCCACACTGCTGACAGCGCCTGGCTGACCAAGACGTGGTCCCGAG	240
Db	181	GCGCTGGACAGCGCCTTCCACACTGCTGACAGCGCCTGGCTGACCAAGACGTGGTCCCGAG	240
QY	241	GACAAGTTTATAGGAGAGCGTTCATCTGAAAGGAAAGAGGCGTGGCCCGACGGCGTTCCAG	300
Db	241	GACAAGTTTATAGGAGAGCGTTCATCTGAAAGGAAAGAGGCGTGGCCCGACGGCGTTCCAG	300
QY	301	GCCCTCTGTCTCTGGCTGCTAGCCACGAGACTCCACAGCCATCCTGGACTTCTGGAAGGTG	360
Db	301	GCCCTCTGTCTCTGGCTGCTAGCCACGAGACTCCACAGCCATCCTGGACTTCTGGAAGGTG	360
QY	361	CTGTTCGAAGGACTACAACTCGAAGCGCTATGGCGGCTGACGCCATCTCTGACAGACTTC	420
Db	361	CTGTTCGAAGGACTACAACTCGAAGCGCTATGGCGGCTGACGCCATCTCTGGAAGACTTC	420
QY	421	CCCAAGATGTGGACTCTCAGCCAGCGCCCGGAAGGGAGAGACCCCGCGCGTCCCAAG	480
Db	421	CCCAAGATGTGGACTCTCAGCCAGCGCCCGGAAGGGAGAGACCCCGCGCGTCCCAAG	480
QY	481	GCTTTGCTACCGCCACCCAGACTCCCAACAGAGAGAGGCTCTGAAAGAGCTCGAGCT	540
Db	481	GCTTTGCTACCGCCACCCAGACTCCCAACAGAGAGAGGCTCTGAAAGAGCTCGAGCT	540
QY	541	GCGGGCCGAGAGCCCTGACTCCAAAGGGGACCGCGCAGCCAGGCTCTCACTGAAGGCC	600
Db	541	GCGGGCCGAGAGCCCTGACTCCAAAGGGGACCGCGCAGCCAGGCTCTCACTGAAGGCC	600

XX AX58606;
 AC 16-AUG-1999 (first entry)
 XX
 DE Mouse autoimmune polyglandular disease type 1 (APGD1) cDNA.
 XX
 KM Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;
 KM APECED: autoimmune polyglandular disease type 1; APGD1: AIRE;
 KM transcription factor; autoimmune disease; diagnosis; gene therapy;
 KM mouse; animal model; ss.
 XX
 OS Mus sp.
 XX WO918197-A2.
 XX
 PD 15-APR-1999.
 XX
 PF 02-OCT-1998; 98WO-EP06294.
 XX
 PR 12-NOV-1997; 97EP-0119810.
 PR 02-OCT-1997; 97EP-0117154.
 PR 08-OCT-1997; 97EP-0117398.
 XX
 PA (PLAC) MAX PLANCK GES. FORDERUNG WISSENSCHAFTEN.
 PA (NAPU-) NAT. PUBLIC HEALTH INST.
 PI Aaltonen J, Björnsen P, Horelli-Kuitunen N, Lehnach H;
 PI Palotie A, Peltonen L, Perheentupa J, Yaspo M;
 DR P-PSDB: AAY06074.
 DR WPI: 1999-287735/24.
 PT New polypeptide which co-segregates in mutated form - with
 PT autoimmune polyendocrinopathy candidiasis ectodermal dystrophy
 XX
 PS Claim 6; Fig 14A-C; 77pp; English.
 XX
 CC This is the nucleotide sequence of murine AIRE cDNA. Murine AIRE
 CC is the homologue of the human AIRE (or autoimmune polyglandular
 CC disease type 1, APGD1) gene (see AX58605). The overall nucleotide
 CC sequence identity between the mouse AIRE coding sequence and that
 CC of the human is 76.67%. The mAIRE gene was localised to chromosome
 CC 10. Human mutated APGD1 co-segregates with autoimmune
 CC polyendocrinopathy candidiasis ectodermal dystrophy (APECED). The
 CC murine homologue may be used to develop an animal model for APECED,
 CC to examine the events that lead to the development of APECED and
 CC possibly to develop agents for preventing and/or treating this
 CC autoimmune disease.
 XX
 SQ Sequence 1656 BP; 344 A; 532 C; 479 G; 301 T; 0 other;
 Query Match 42.9%; Score 964.2; DB 20; Length 1656;
 Best Local Similarity 76.0%; Pred. No. 3.3e-179;
 Matches 1249; Conservative 0; Mismatches 373; Indels 21; Gaps 4;
 129 GAGACGGCGGCGCTGAGCGGCTTCGAGCGTCACACCGACAGAGATCGCGGTGCGGTGA 188
 DB 12 GGTGTGAATGCTACGCGGTGTGCTGAGGCTGCACCGACGAGATCGGGGTGGCCATAGA 71
 QY 189 CAGCGCGTTCCTCCACTGCTGCACGCGGTGCTGACACGAGCTGTCCCGGAGCAAGTT 248
 DB 72 CAGTGCCTTTCCGCTGTGATGCTTACGCGACACGAGCTGTCCCTGAGGACAAATT 131
 QY 249 TAGAGGAGCGCTTCACTGTGAAGAAAGAGGGCTGCCCGAGGCTTCCACGCGCTCT 308
 DB 132 CAGAGGAGCGCTCTGCTGTAAGAGAAAGAGGCTGCCCGAGGCTTCCACGCGCTGCT 191
 QY 309 GTCCTGGCTGCTGACCGACGAGCTCCACAGCATCTTGGACTTCTGGAGGGTCTGTTCAA 368
 DB 192 GTCTGTGGCTCTGACCGCGGACAGTGGGGCATCTGGAATTTCTGGAGGATTTCTTTAA 251
 QY 369 GGACTACAACTGAGCGCTATGGCGGCTGTCAGGCCATCTGAGACGATTTCCCAAGA 428

DB 252 GGACTACAACTGAGCGGTACACCGCGCTCTCATAGCATCTGAGCGGCTTCCCAAGA 311
 QY 429 TGTGACCTAGCGACGCGCGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
 DB 312 TGTGACCTAGCGACGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 371
 QY 489 ACCGCCACCCAGACTCCCGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 548
 DB 372 ACTGCCACCCAG 431
 QY 549 AGCAGCCCTGACTCCAG 608
 DB 432 AGCACTCTGCTCTCAAG 491
 QY 609 CAAGAGCCGAG 668
 DB 492 TAAG 551
 QY 669 CATGTACGCTTCACTGACAG 728
 DB 552 CATGTACGCTTCACTGACAG 611
 QY 729 AGGGGCGGTGAG 788
 DB 612 AGGGGCGGTGAG 671
 QY 789 CATTCAGAGTGTGGAG 845
 DB 672 CATTCAGAGTGTGGAG 731
 QY 846 GAAGAAAG 905
 DB 732 GAAGAAAG 791
 QY 906 GGGGCGGTGAG 965
 DB 792 GGTCTCATATACGTGTGAG 851
 QY 966 TCTGGCCCTCCCGAGTACAG 1025
 DB 852 TCTGGCCCTCCCGAGTACAG 911
 QY 1026 TCGGAGCGCGGAG 1085
 DB 912 CCACGAGCGGAG 971
 QY 1086 CCGTCCCGCTCCCGAGTACAG 1145
 DB 972 CCGTCCCGCTCCCGAGTACAG 1031
 QY 1146 GGCAGAGTCCAG 1205
 DB 1032 GGGAGAGAGTCCAG 1091
 QY 1206 GGAAGAGCGCGTCCCGAGTACAG 1265
 DB 1092 AGAGAGCGCGTCCCGAGTACAG 1151
 QY 1266 GGAAGAGCGCGTCCCGAGTACAG 1325
 DB 1152 GGAAGAGCGCGTCCCGAGTACAG 1211
 QY 1326 TGCAGAGCGCGTCCCGAGTACAG 1385
 DB 1212 TGCAGAGCGCGTCCCGAGTACAG 1265
 QY 1386 TGAAGGTACAG 1445
 DB 1266 TGAAGGTACAG 1325
 QY 1446 CGTGTGGGTACTCACTGCGCGGCTGCTTCCACTGGCGTGCCTTCCAGCCGG 1505

Qy	2116	TCGCGCCACTGCATCCAGTCTGCTCGCAAGAGTGAGACTCCGTCGCAAAACAAACA	2175
Db	1744	TTGCGCCACTGCATCCAGCTGGGTGACAAAGAGTCAATCCATTCGCAAAAAAAAAA	1685
Qy	2176	AAAAACCATACATATAA 2195	
Db	1684	AAAAAAAAAGAAAAAAAAGA 1665	
RESULT 8			
AAK68127			
XX	AAK68127	standard; DNA; 5540 BP.	
XX	AC	AAK68127;	
XX	DT	06-NOV-2001 (first entry)	
XX	XX		
XX	XX	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22939.	
XX	XX	Human: immune; haematopoietic; immune/haematopoietic antigen; cancer;	
XX	XX	cytostatic; gene therapy; vaccine; metastasis; ds.	
XX	OS	Homo sapiens.	
XX	PN	WO200157182-A2.	
XX	PD	09-AUG-2001.	
XX	PF	17-JAN-2001; 2001WO-US01354.	
XX	PR	31-JAN-2000; 2000US-0179065.	
XX	PR	04-FEB-2000; 2000US-0180628.	
XX	PR	24-FEB-2000; 2000US-0184664.	
XX	PR	02-MAR-2000; 2000US-0186350.	
XX	PR	16-MAR-2000; 2000US-0189874.	
XX	PR	17-MAR-2000; 2000US-0190076.	
XX	PR	18-APR-2000; 2000US-0198123.	
XX	PR	19-MAY-2000; 2000US-0205515.	
XX	PR	07-JUN-2000; 2000US-0209467.	
XX	PR	28-JUN-2000; 2000US-0214886.	
XX	PR	30-JUN-2000; 2000US-0215135.	
XX	PR	07-JUL-2000; 2000US-0216647.	
XX	PR	11-JUL-2000; 2000US-0216880.	
XX	PR	11-JUL-2000; 2000US-0217487.	
XX	PR	11-JUL-2000; 2000US-0217496.	
XX	PR	14-JUL-2000; 2000US-0218290.	
XX	PR	26-JUL-2000; 2000US-0220963.	
XX	PR	14-AUG-2000; 2000US-0220964.	
XX	PR	14-AUG-2000; 2000US-0224518.	
XX	PR	14-AUG-2000; 2000US-0224519.	
XX	PR	14-AUG-2000; 2000US-0225213.	
XX	PR	14-AUG-2000; 2000US-0225214.	
XX	PR	14-AUG-2000; 2000US-0225266.	
XX	PR	14-AUG-2000; 2000US-0225267.	
XX	PR	14-AUG-2000; 2000US-0225268.	
XX	PR	14-AUG-2000; 2000US-0225270.	
XX	PR	14-AUG-2000; 2000US-0225447.	
XX	PR	14-AUG-2000; 2000US-0225757.	
XX	PR	14-AUG-2000; 2000US-0225758.	
XX	PR	14-AUG-2000; 2000US-0225759.	
XX	PR	18-AUG-2000; 2000US-0226279.	
XX	PR	22-AUG-2000; 2000US-0226281.	
XX	PR	22-AUG-2000; 2000US-0226681.	
XX	PR	22-AUG-2000; 2000US-0226868.	
XX	PR	23-AUG-2000; 2000US-0227009.	
XX	PR	30-AUG-2000; 2000US-0228924.	
XX	PR	01-SEP-2000; 2000US-0229287.	
XX	PR	01-SEP-2000; 2000US-0229343.	
XX	PR	01-SEP-2000; 2000US-0229344.	
XX	PR	01-SEP-2000; 2000US-0229345.	
XX	PR	05-SEP-2000; 2000US-0229509.	
XX	PR	05-SEP-2000; 2000US-0229513.	
XX	PR	06-SEP-2000; 2000US-0230437.	

[illegible]

XX DE Human genomic DNA for PHIP/NDRP.
 XX KW PHIP; pleckstrin homology domain-interacting protein; NDRP; ds;
 KW neuronal differentiation-related protein; insulin receptor substrate;
 KW IRS; signal transducer and activator of transcription; STAT;
 KW transgenic animal; diabetes mellitus type 2; hyperglycaemia;
 KW myotonic muscular dystrophy; acanthosis; nigricans; retinopathy;
 KW nephropathy; arteriosclerosis; peripheral arterial disease; cancer;
 KW adenocarcinoma; leukaemia; breast cancer; prostate cancer; colon cancer;
 KW ovarian cancer; autoimmune disease; inflammation; immunodeficiency.
 XX OS Homo sapiens.
 XX PN WO200185785-A2.
 XX PD 15-NOV-2001.
 XX PF 10-MAY-2001; 2001WO-CA00673.
 XX PR 11-MAY-2000; 2000US-203561P.
 XX PA (ROZA/) ROZAKIS-ADCOCK M.
 XX PA (FARH/) FARHANG-FALLAH J.
 XX PA (CHEN/) CHENG A.
 XX PI Rozakis-Adcock M, Farhang-Fallah J, Cheng A;
 XX DR WPI; 2002-041586/05.
 XX PT Novel Pleckstrin homology domain interacting protein recruiting
 PT proteins of insulin receptor substrate family, and signal transducer
 PT and activator of transcription factors to their receptors, useful to
 PT treat diabetes
 PS Disclosure: Page 99-133; 139pp; English.
 XX CC The invention relates to an isolated pleckstrin homology domain
 CC interacting protein (PHIP) that recruits proteins of the insulin
 CC receptor substrate (IRS) family, and signal transducer and activator of
 CC transcription (STAT) transcription factors, to receptors that interact
 CC with and phosphorylate the proteins and STAT transcription factors.
 CC the nucleic acid encoding PHIP (nPHIP), a nucleic acid which binds to
 CC nPHIP or regions of it, analogues, fragments or allelic variants of PHIP
 CC or nPHIP, a nucleic acid sequence having substantial sequence identity or
 CC sequence similarity with a nucleic acid sequence fully defined
 CC human neuronal differentiation-related protein (NDRP) nucleic acid
 CC sequence or its exons as given in the specification, expression
 CC vectors and host cells expressing the nucleic acids, anti-PHP antibodies,
 CC and a transgenic animal not already expressing PHIP. The nucleic
 CC acids, proteins and antibodies are useful for diagnosis and treatment of
 CC a condition associated with an insulin receptor (e.g. diabetes mellitus
 CC type 2, hyperglycaemia, myotonic muscular dystrophy, acanthosis,
 CC arterial disease), retinopathy, nephropathy, arteriosclerosis, peripheral
 CC cancer, prostate cancer, colon cancer, ovarian cancer and many others
 CC given in the specification), autoimmune disease, inflammation and
 CC immunodeficiency. The protein is also useful for discovering
 CC or testing compounds which may be either enhancers or inhibitors of PHIP
 CC function. The present sequence is genomic DNA encoding PHIP and NDRP.
 XX
 SO Sequence 140036 BP; 48500 A; 25843 C; 22499 G; 43194 T; 0 other;
 Query Match 6.8%; Score 152; DB 24; Length 140036;
 Best Local Similarity 79.9%; Pred. No. 2.5e-20;
 Matches 179; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

DB 1994 AGGGGGCTAGAGCAGAGAAATCTTTGAACCCAGGAGGGGAGTTGCAGTGAAGTGA 2053
 QY 2116 TTGCGCCACTGACACTGCTGCTGGCAAGAGTGAAGTCTCGTCAAAAACAAACA 2175
 DB 2054 TCACGACATGACATCCAGCTGGGCGACAAAGAGGAGACTTCATCTCAAAAACAAACA 2113
 QY 2176 AAAAACACATATACATTAATTTATCATCTGACACCTTTTACG 2219
 DB 2114 AAACAAACAAACAAACACATGTTAAGAGACTTACCAAGGTCAG 2157
 RESULT 11
 ABA20762/C
 ID ABA20762 standard; DNA; 26427 BP.
 XX AC ABA20762;
 XX DT 23-JUN-2002 (first entry)
 XX DE Human nervous system related polynucleotide SEQ ID NO 13093.
 KW Human; neurotropic; neuroprotective; cytosolic; dermatological; virocidic;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnereary;
 KW antiparkinsonian; antislacking; antianaemic; antiarthritic; cancer;
 KW antineumatic; hepatotropic; cerebrotropic; antiinflammatory;
 KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
 KW antiparasitic; cardiatic; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
 XX OS Homo sapiens.
 XX PN WO200159063-A2.
 XX PD 16-AUG-2001.
 XX PT 17-JUN-2001; 2001WO-US01334.
 XX PF 31-JAN-2000; 2000US-0179065.
 XX PR 04-FEB-2000; 2000US-0180628.
 XX PR 24-FEB-2000; 2000US-0184664.
 XX PR 02-MAR-2000; 2000US-0186350.
 XX PR 16-MAR-2000; 2000US-0189874.
 XX PR 17-MAR-2000; 2000US-0190076.
 XX PR 18-APR-2000; 2000US-0198123.
 XX PR 19-MAY-2000; 2000US-0205515.
 XX PR 07-JUN-2000; 2000US-0209467.
 XX PR 28-JUN-2000; 2000US-0214886.
 XX PR 30-JUN-2000; 2000US-0215135.
 XX PR 07-JUL-2000; 2000US-0216647.
 XX PR 11-JUL-2000; 2000US-0216880.
 XX PR 11-JUL-2000; 2000US-0217487.
 XX PR 14-JUL-2000; 2000US-0217496.
 XX PR 14-JUL-2000; 2000US-0218290.
 XX PR 26-JUL-2000; 2000US-0220963.
 XX PR 26-JUL-2000; 2000US-0220964.
 XX PR 14-AUG-2000; 2000US-0224518.
 XX PR 14-AUG-2000; 2000US-0224519.
 XX PR 14-AUG-2000; 2000US-0225213.
 XX PR 14-AUG-2000; 2000US-0225214.
 XX PR 14-AUG-2000; 2000US-0225266.
 XX PR 14-AUG-2000; 2000US-0225267.
 XX PR 14-AUG-2000; 2000US-0225268.
 XX PR 14-AUG-2000; 2000US-0225270.
 XX PR 14-AUG-2000; 2000US-0225447.
 XX PR 14-AUG-2000; 2000US-0225757.
 XX PR 14-AUG-2000; 2000US-0225758.
 XX PR 14-AUG-2000; 2000US-0225759.
 XX PR 18-AUG-2000; 2000US-0226279.
 XX PR 22-AUG-2000; 2000US-0226681.
 XX PR 22-AUG-2000; 2000US-0226686.
 XX PR 22-AUG-2000; 2000US-0227182.
 XX PR 23-AUG-2000; 2000US-0227009.
 XX PR 30-AUG-2000; 2000US-0228924.


```

PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251130.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2000US-0254097.
PR 05-JAN-2001; 2000US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases.
XX
XX Disclosure; SEQ ID NO 13094; 1701bp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins
XX (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and antagonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 26427 BP; 5634 A; 7262 C; 7074 G; 6457 T; 0 other;
XX
XX Query Match 6.8%; Score 151.6; DB 22; Length 26427;
XX Best Local Similarity 82.3%; Pred. No. 2.4e-20;
XX Matches 186; Conservative 0; Mismatches 39; Indels 1; Gaps 1;
XX
XX 1996 CTACAAAAATATTAATTTAGCTGGTGTGCTGGTGCCTGTAAATCCAGCTACAT 2055
XX DB 4490 CTACTAAAAATATTAATTTAGCTGGTGTGCTGGTGCCTGTAAATTCAGCTACAT 4431
XX
XX 2056 GGGAGCCTGAGCATGAGAAATCACTGAGCTGGGAGGAGGAGCTTGACAGCTGAGA 2115
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX DB 4430 GGGAGCTGAGGAGGAGAAATGCTTGAACCGGGAGGAGGAGGAGCTTGACAGCTGAGA 4371
XX
XX 2116 TTGGCCCACTGACATCCAGTGTGCTGGCAAGAGTGAGACTCGCTCAAAAACAAACA 2175
XX -||| -||| -||| -||| -||| -||| -||| -||| -||| -||| -||| -||| -|||

```

```

DB 4370 TTGCACCACTGCATCCACCTGGGCAAC-AGAGCGAGACTCGCTCAAAAAAAAAA 4312
OY 2176 AAAAAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2221
DB 4311 AAAAAAAAAAAGAGTTACTTTAATTCCTACCTCTTACTTT 4266
XX
XX RESULT 13
XX ABQ88186
XX ID ABQ88186 standard; cDNA; 169739 BP.
XX
XX ABQ88186;
XX
XX 18-SEP-2002 (first entry)
XX
XX Human osteoblast differentiation related cDNA SEQ ID NO 93.
XX
XX Human; osteoblast; stem cell differentiation; bone tissue deposition;
XX osteoporosis; osteopathic; ss.
XX
XX Homo sapiens.
XX
XX WO200250301-A2.
XX
XX 27-JUN-2002.
XX
XX 18-DEC-2001; 2001WO-US48276.
XX
XX 18-DEC-2000; 2000US-255882P.
XX 24-APR-2001; 2001US-285691P.
XX
XX (GENE-) GENE LOGIC INC.
XX (PROC) PROCTER & GAMBLE CO.
XX
XX Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
XX Wertz L;
XX WPI; 2002-557663/59.
XX
XX Use of genes and their expression profiles associated with osteoblast
XX differentiation for screening modulators bone formation, for diagnosing
XX or treating e.g. osteoporosis, or as markers for the differentiation
XX process.
XX
XX Claim 1; SEQ ID NO 93; 78bp + Sequence Listing; English.
XX
XX The invention relates to genes and their expression profiles are used
XX for:
XX (a) screening modulators of precursor stem cell differentiation into
XX osteoblasts, or bone tissue deposition;
XX (b) diagnosing abnormal deposition of bone tissue, abnormal rate of
XX osteoblast formation or osteoporosis; or
XX (c) treating or monitoring treatment of the conditions cited in (b), or
XX monitoring the progression of bone tissue deposition.
XX Specific conditions include postmenopausal osteoporosis, glucocorticoid
XX drug-induced abnormalities in bone formation or bone loss, conditions
XX that involve altered bone metabolism (e.g. idiopathic juvenile
XX osteoporosis), skeletal disease linked to breast cancer, mastocytosis,
XX Fanconi syndrome or fibrous dysplasia. The present sequence is that of an
XX osteoblast differentiation associated cDNA marker of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 169739 BP; 49809 A; 35660 C; 35715 G; 48555 T; 0 other;
XX
XX Query Match 6.8%; Score 151.6; DB 24; Length 169739;
XX Best Local Similarity 85.4%; Pred. No. 3.1e-20;
XX Matches 169; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
XX
XX 1996 CTACTAAAAATATTAATTTAGCTGGTGTGCTGGTGCCTGTAAATCCAGCTACAT 2055
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

QY 2056 GGGAGCCTGAGCATGAGAACTTGAACCTCGGAGGTGACGTGAGCTGAGA 2115
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 393 AGGAGGCTGAGCAGGAACTGCTTGAAACCGGAGCAGAGGTGACGTGAACCTTAA 334
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2116 TTGGCGCACTGCACTGCTGCTGGGCAAGAGTGAGACTCGCTCAAAACAAACA 2175
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 333 TTGCGCCATTGCACTCCAGCCTGGGTGACAGAGTGAACCTCGCTCAAAACAAACA 274
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2176 AAAAAACCATATACATTAATTTATCATCTCGACCACTTTTCAGTTCAGTGCATTCACA 2235
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 273 AAAAAACCATATACATTAATTTATCATCTCGACCACTTTTCAGTTCAGTGCATTCACA 214
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2236 TCTCATGTAA 2245
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 213 ATTCTGGCAA 204
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: February 14, 2003, 12:47:08
 Job time : 961.856 secs

[illegible]

Query Match	6.7%	Score 150.2	DB 9	Length 3535
Best Local Similarity	80.4%	Pred. No. 3e-21		
Matches 176; Conservative	0	Mismatches 43	Indels 0	Gaps 0
1996 CTACTAAATAATMAAATTAGCTGGGTGTGTGTGGTCCCTGTATATCCACGTACAT	2055			

PRIOR APPLICATION NUMBER: US01/01360	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,867	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01344	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,892	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01345	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,888	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01329	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,905	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01354	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,891	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01339	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,869	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01340	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,874	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01334	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,898	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01320	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,853	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01349	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,902	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01239	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,870	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01348	PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,882	PRIOR FILING DATE: 2001-01-17

```

? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: US01/01347
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: 09/764, 896
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: US01/01307
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: 09/764, 864
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: US01/01341
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: 09/764, 856
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: US01/01336
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: 09/764, 868
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: US01/01312
? PRIOR FILING DATE: 2001-01-17
? PRIOR APPLICATION NUMBER: 60/179, 065
? PRIOR FILING DATE: 2000-01-31
? PRIOR APPLICATION NUMBER: 60/180, 628
? PRIOR FILING DATE: 2000-02-04
? PRIOR APPLICATION NUMBER: 60/209, 467
? PRIOR FILING DATE: 2000-06-07
? NUMBER OF SEQ ID NOS: 167
? SOFTWARE: Patentln Ver. 2.0
? SEQ ID NO 160
? LENGTH: 1186
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: SITE
? LOCATION: (342)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (343)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (344)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (345)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (346)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (347)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (348)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (349)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (350)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (351)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (352)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (353)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (354)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (355)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (356)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (357)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (358)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (359)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (360)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (361)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (362)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (363)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (364)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (365)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (366)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (367)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (368)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (369)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (370)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (371)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (372)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (373)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (374)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (375)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (376)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (377)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (378)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: SITE
? LOCATION: (379)
? OTHER INFORMATION: n equals a,t,g, or c

```

```

1 LOCATION: 6058..6076
2 OTHER INFORMATION: 99-16073.rp
3 NAME/KEY: primer_bind
4 LOCATION: 6493..6512
5 OTHER INFORMATION: 99-16073.pu complement
6 NAME/KEY: primer_bind
7 LOCATION: 9661..9678
8 OTHER INFORMATION: 99-16074.rp
9 NAME/KEY: primer_bind
10 LOCATION: 10119..10136
11 OTHER INFORMATION: 99-16074.pu complement
12 NAME/KEY: primer_bind
13 LOCATION: 14754..14774
14 OTHER INFORMATION: 99-13817.pu
15 NAME/KEY: primer_bind
16 LOCATION: 15183..15203
17 OTHER INFORMATION: 99-13817.rp complement
18 NAME/KEY: primer_bind
19 LOCATION: 20137..20155
20 OTHER INFORMATION: 99-16066.pu
21 NAME/KEY: primer_bind
22 LOCATION: 20569..20588
23 OTHER INFORMATION: 99-16066.rp complement
24 NAME/KEY: misc_binding
25 LOCATION: 3239..3285
26 OTHER INFORMATION: 99-16063-218.probe
27 NAME/KEY: misc_binding
28 LOCATION: 6208..6254
29 OTHER INFORMATION: 99-16073-282.probe
30 NAME/KEY: misc_binding
31 LOCATION: 9848..9894
32 OTHER INFORMATION: 99-16074-266.probe
33 NAME/KEY: misc_binding
34 LOCATION: 14945..14991
35 OTHER INFORMATION: 99-13817-215.probe
36 NAME/KEY: misc_binding
37 LOCATION: 20236..20282
38 OTHER INFORMATION: 99-16066-123.probe
39 NAME/KEY: primer_bind
40 LOCATION: 3243..3261
41 OTHER INFORMATION: 99-16063-218.mis
42 NAME/KEY: primer_bind
43 LOCATION: 3263..3281
44 OTHER INFORMATION: 99-16063-218.mis complement
45 NAME/KEY: primer_bind
46 LOCATION: 6212..6230
47 OTHER INFORMATION: 99-16073-282.mis
48 NAME/KEY: primer_bind
49 LOCATION: 6232..6250
50 OTHER INFORMATION: 99-16073-282.mis complement
51 NAME/KEY: primer_bind
52 LOCATION: 9852..9870
53 OTHER INFORMATION: 99-16074-266.mis
54 NAME/KEY: primer_bind
55 LOCATION: 9872..9890
56 OTHER INFORMATION: 99-16074-266.mis complement
57 NAME/KEY: primer_bind
58 LOCATION: 14949..14967
59 OTHER INFORMATION: 99-13817-215.mis
60 NAME/KEY: primer_bind
61 LOCATION: 14969..14987
62 OTHER INFORMATION: 99-13817-215.mis complement
63 NAME/KEY: primer_bind
64 LOCATION: 20240..20258
65 OTHER INFORMATION: 99-16066-123.mis
66 NAME/KEY: primer_bind
67 LOCATION: 20260..20278
68 OTHER INFORMATION: 99-16066-123.mis complement
69 NAME/KEY: misc.feature
70 LOCATION: 285.982.2264.2273.2823.11227.11232.13663.13681.13842..13843
71 OTHER INFORMATION: n=a, g, c or t
72 US-09-416-384A-2

```



```

Db 13454512 CTACATAAAAAATTGTGAAAATACTGGTGTTGGTGGCGGGCCGCTGTAATGCCAGCTACTC 13453533
OY      2056 GGGAGCCTGAGGCATGAGAATCACTTGAAC TCGGAGGTGAGGTTGCAGTAGCTGAGA 2115
          ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13453532 AGGAGGCTGAGGACGAGAGAAATCACTTGAAC CCGGAGGACGACAGTGTTCAGTAGCTGAGA 13452933
OY      2116 TTGCGCGCATCTGCACGTCTGGTGGCGGAAGAGTAGAGACTCCGTCGCAAAAACAACA 2175
          | | ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 13452932 TCGTGCCATTGCACTTCAGCCTGGGAGACAMAGCAAGACTGTCTCAAAAAAAAAAAAAA 13452333
OY      2176 AAAAAACCAATATA 2188
          ||||| ||| |
Db 13452332 AAAAAAAAAAAAAA 1345220

RESULT 14
US-09-795-686-1/c
; Sequence 1, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreln
; APPLICANT: Steinhorsdottir, Valgerdur
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345_2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; PRIORITY FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) ..(1531)
; OTHER INFORMATION: r=g or a
; NAME/KEY: misc_feature
; LOCATION: (1) ..(1531)
; OTHER INFORMATION: y=t/u or c
; NAME/KEY: misc_feature
; LOCATION: (1) ..(1531)
; OTHER INFORMATION: m=a or c
; NAME/KEY: misc_feature
; LOCATION: (1) ..(1531)
; OTHER INFORMATION: s=g or c
; NAME/KEY: misc_feature
; LOCATION: (1) ..(1531)
; OTHER INFORMATION: w=a or t/u
; NAME/KEY: misc_feature
; LOCATION: (1) ..(1531)
; OTHER INFORMATION: b=g or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1) ..(1531)
; OTHER INFORMATION: d=a or g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1) ..(1531)
; OTHER INFORMATION: h=a or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1) ..(1531)
; OTHER INFORMATION: v=a or g or c
; NAME/KEY: misc_feature
; LOCATION: (1) ..(1531)
; OTHER INFORMATION: n=a or g or c or t/u
US-09-795-686-1

Query Match      6.5%; Score 146.6; DB 10; Length 1503841;
Best Local Similarity 85.0%; Pred. No. 5,1e-20;
```

```

Matches 164; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 1996 CTACTAAAAATATAAAAATTAGCTGGGTGTGGTGGCTGCTTAATCCAGCTACAT 2055
      |||||
Db 1345412 CTACTAAAAATATGAAAAATAGCTGGGTGTGGTGGCTGCTTAATCCAGCTACTC 1345353
QY 2056 GGGAGCGCTGAGCATGAGATCACTTGAACCTGGGAGGTGGAGGTTCAGCTGAGCTGAGA 2115
      |||||
Db 1345352 AGGAGCGCTGAGGCGAGGAGATATCTTGAACCCGGGAGGCAAGGTTGCAATGAGCTGAGA 1345293
QY 2116 TTGGCGCAGCTGCACCTCCAGCTGTGCTGGCAGAGAGTGAAGCTCCGTCCTCAAAAACAACA 2175
      |||||
Db 1345292 TCGTGCACCTTGACCTCCAGCCTGGGAGACAGACAGACTCTGTCTCAAAAAA 1345233
QY 2176 AAAAAACCATATCA 2188
      |||||
Db 1345232 AAAAAAAAAAGA 1345220

RESULT 15
US-09-818-512-3
; Sequence 3, Application US/09818512
; Patent No. US20020142416A1
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: C1001192
; CURRENT APPLICATION NUMBER: US/09/818, 512
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 116592
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(116592)
; OTHER INFORMATION: n = A,T,C or G
US-09-818-512-3

Query Match 6.5%; Score 146.4; DB 10; Length 116592;
Best Local Similarity 84.2%; Pred. No. 3,4e-20;
Matches 165; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 1996 CTACTAAAAATATAAAAATTAGCTGGGTGTGGTGGCTGCTTAATCCAGCTACAT 2055
      |||||
Db 39599 CTACTAAAAATATCAAAAAATTAGCTGGGTGTGGTGAAGTGGCTGTAATCCAGCTACATC 39658
QY 2056 GGGAGCGCTGAGGCGATGAGATCACTTGAACCTGGGAGGTGGAGGTGAGCTGAGCTGAGA 2115
      |||||
Db 39659 AGGAGCGCTGAGGCGGAGGAAATTGCTTTGAGCCCAATGAGGTGAGCTGAGTGTGAGAGA 39718.
QY 2116 TTGGCGCAGCTGCACCTCCAGCTGTGCTGGCAGAGTGAAGCTCCGCTCAAAAAACAANA 2175
      |||||
Db 39719 TTGGCGCAGCTGCACCTCCAGCTGGGAGCAACAAGATGTGAAGCTCATCTCAAAAAA 39778
QY 2176 AAAAAACCATATCA 2191
      |||||
Db 39779 AAAAAAAAAA 39794

```

Search completed: February 14, 2003, 19:51:07
Job time : 2394.54 secs

GenCore version 5.1.4_P5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:51:42 ; Search time 21 Seconds

(without alignments)
2494.917 Million cell updates/sec

Title: US-09-509-595b-2

Perfect score: 2902
Sequence: 1 MATDAILRLLRLHREIAV.....DGLQWAIQSMARPAAPPS 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	273.5	9.4	530	2 I38558	Mi-2 autoantigen 2
2	221.5	7.6	3938	2 T42761	Basoon protein -
3	219.5	7.6	1829	2 T34239	hypothetical prote
4	219	7.5	371	2 B49515	phosphoprotein 75
5	217	7.5	1051	2 S55259	TIF1 protein - mou
6	206	7.1	1787	2 T20160	hypothetical prote
7	194.5	6.7	4957	2 T03455	ALR protein - huma
8	194.5	6.4	5262	2 T03454	hypothetical prote
9	186.5	6.4	607	2 S58151	hypothetical prote
10	179.5	6.2	404	2 S62475	PHD finger protein
11	178.5	6.2	3942	2 T42730	Basoon protein -
12	171	5.9	825	2 T02518	hypothetical prote
13	168.5	5.8	660	1 Q00B3	BHLFI protein - hu
14	168	5.8	1237	2 T45070	protein kinase hom
15	161	5.5	684	2 S52835	hypothetical prote
16	160	5.5	350	2 G01950	hypothetical prote
17	159	5.5	958	2 H84783	probable PHD-type
18	157.5	5.4	1201	2 T29329	hypothetical prote
19	155	5.3	839	2 T45908	hypothetical prote
20	154	5.3	1146	2 A38587	collagen, cornea-s
21	153	5.3	1006	2 T42731	atrophin-1 related
22	152.5	5.3	813	2 E85135	hypothetical prote
23	152.5	5.3	839	2 F75518	hypothetical prote
24	152	5.2	810	2 G01252	small GTP binding
25	152	5.2	1049	1 CGB075	collagen alpha 1(I
26	152	5.2	1204	2 T19918	hypothetical prote
27	152	5.2	1736	2 T00391	hypothetical prote
28	151.5	5.2	728	2 S57142	hypothetical prote
29	151.5	5.2	1479	2 T17401	transcription regu

30	151.5	5.2	1532	2 A61262	collagen alpha 1(X
31	150.5	5.2	801	2 T29018	hypothetical prote
32	150	5.2	2274	2 T30258	adenomatous polyo
33	149.5	5.2	1151	2 T18535	high molecular mas
34	149.5	5.2	1320	2 JC5630	TCOF1 protein - mo
35	149	5.1	1207	2 T00378	KIAA0641 protein -
36	149	5.1	2187	2 T30826	nascent polypeptid
37	148.5	5.1	283	2 T11644	hypothetical prote
38	148.5	5.1	1400	2 T31555	hypothetical prote
39	148	5.1	640	2 T08179	LRG5 protein - chl
40	147	5.1	574	2 T43556	Wiskott-Aldrich sy
41	147	5.1	1093	2 T38533	AF17 protein - hum
42	146	5.0	574	2 T38533	Wiskott-Aldrich sy
43	144.5	5.0	1214	2 JC2069	zinc-finger protei
44	144	5.0	455	2 A87913	protein B0205.10 f
45	144	5.0	649	2 T46500	hypothetical prote

ALIGNMENTS

RESULT 1

I38558

Mi-2 autoantigen 240 kDa protein - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 05-Nov-1999

C:Accession: I38558

R:Ge, O.; Nilasena, D.S.; O'Brien, C.A.; Frank, M.B.; Taroff, I.N.

J. Clin. Invest. 96, 1730-1737, 1995

A:Title: Molecular analysis of a major antigenic region of the 240-kD protein of Mi-2

A:Reference number: I38558; MUID:96013633; PMID:7560064

A:Accession: I38558

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-530 <RES>

A:Cross-references: EMBL:008379; NID:9761717; PIDN:AAC50228.1; PID:9761718

Query Match 9.4%; Score 273.5; DB 2; Length 530;

Best Local Similarity 23.6%; Pred. No. 7.1e-09;

Matches 102; Conservative 43; Mismatches 160; Indels 127; Gaps 15;

QY	116	PAVPKALVPPRL-----PIRKASEEARAAPALPRGTASPSQILAKPPKPESSA	170
DB	116	PSGPPLPPLPPADIDPPPIRAKTKGK-----GPHKRRKSPRPD----	159
QY	171	EQGRPLPGNGIQTMASAVQRAVAMSSGDVPGARGAVESILLQVFPESGSKCIQVGGEE	230
DB	160	-----GRKKLRGKKMAPLKITIKILGGR-----KKGSS-----DEGPEP	195
QY	231	YTPSKFEDSGSGKKAKRSSGPKPLVRAKGAOGAAPPGGGEAR-LGQGSVPAPLALPSDP	289
DB	196	EAEESDLDSSSVHSASGRPDG---VTKTKLRRPRGKKKKVUG-----CPAVGEEVD	248
QY	290	QLHOKNEDECAVCRDGGELICDCCPRAPHLACSPPLREIPSGTWKSSCIQATVDEQ	349
DB	249	GVTEDHODYCEVCOQGGELICDCCPRAPHLVCLDPELDRAPEKWKSCPHCEKGVQWEA	308
QY	350	PRAEERPRQPPVETPLPPGLRSAGEEVKRPGRPPGLAGMDTLLYKKHLAPPSAAPLPGL	409
DB	309	KEEEVEEYEE-----GEE-EGEKEEE-----DDHMEY-----	334
QY	410	DSALHLPLLCVGPQGQNLAPGARCGVCGDGTIVLRCTHCAAAFHWCHFPAGTSRPCTG	469
DB	335	-----CRCKDGGELICDCCACISSTHICLNPPLDIPNGE	370
QY	470	LRCRCSGSDVTPAPVEGVL-----APSPARLAPGAPKDDTASHEPALRDLESL	519
DB	371	WLCPRCTCPVLKGRVQKILHMRWCEPVAAPQAQDGNP---DVPPRPRLQGRSEREFF	427
QY	520	LSHETFDGIQW 531	
DB	428	VR---WVGLSYW 436	

RESULT 2

T42761

Bassoon protein - rat

N:Alternate names: brain-specific synapse-associated protein

C:Species: Rattus norvegicus (Norway rat)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000

C:Accession: T42761

R:Dieck, S.; Sammarti-Vila, L.; Langnaese, K.; Richter, K.; Kindler, S.; Soyke, A.; Wex, J. Cell Biol. 142, 499-509, 1998

A:Title: Bassoon, a novel zinc-finger CAG/glutamine-repeat protein selectively localized

A:Reference number: 22249; MUID:98345363; PMID:9679147

A:Accession: T42761

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-3938 <DIE>

A:Cross-references: EMBL:Y16563; NID:93413503; PID:CAN76287.1; PID:93413504

A:Experimental source: strain Sprague Dawley; brain

A:Function:

A:Description: may be involved in cytomatrix organization at the site of neurotransmitter

A:Note: component of the presynaptic cytoskeleton

C:Keywords: coiled coil; zinc finger

Query Match 7.6%; Score 221.5; DB 2; Length 3938; Best Local Similarity 23.4%; Pred. No. 6.1e-05; Matches 123; Conservative 57; Mismatches 232; Indels 113; Gaps 25;

QY 104 VDSQPRKGRPRAPVPRKALVPRPLPTKRKAS--EENARAAPAL-----TPRGTA 153
 DB 225 MDHTAPRSKSOQQLHSPALSPAHSPAKPLGKEPEERSRSPGATQSGPRQAEARATSV 284
 QY 154 PG-----SOLKAPKPKPSSAEQORLPLGNGIQTMSASVQRAVAMSSGGVPGAR 203
 DB 285 PGTQATAPREVGVSPQPLSTKPTAEPRPPAG-----EAQGSATTVPSSL 333
 QY 204 GAVE---GIIIOVFESGGS---KKCIQVGEFYT-----PSK-----FEDSG--SG 242
 DB 334 GAEEOTGGGLTKLFGASLTQASTLMSVQPEADTGGQSPSKGPRKIVFSDAKAG 393
 QY 243 KNAKSSGKPLVRANKAGG--AARGGKRLQGQSVAPRLALPSD-----PQ 290
 DB 394 PRPGSGPGRTPRGATKTERGPRTPGSGPQALAKTGTGTPSKGRADHQAAKAAKPK 453
 QY 291 LHKQNEDECAVC-----RDGGELICDQGP--RAFHACLSPRLRIPGTRCSCC- 340
 DB 454 TPKERAACPLCOALNVSGRGRANVTCTACKLRVCTLCGFNTPHLIVETEMLCINQ 513
 QY 341 ----LQATVQEVQRAEPRPQEVETPLRPGLSAGE---EYKGP--PGEPLAGMDTT 391
 DB 514 TKRLLEGLSGEPAP-LPLPTQEPBPAGVP---QRAAGAPRLKQKQPGQGPQGP----SGS 564
 QY 392 LVYKHLRPPRSAPRLPGLDSSALHRLPLCYVREGQONLAPARCGVCGDGYLRTCHAA 451
 DB 565 LPPKASPOAKAKASPOAKAKASPOAKRPLRASEPSTSSAPRKKTIPIVKAERVPKPPETA 624
 QY 452 AFHWRCHPAGTSRPTGT--RCRSCSGDYTPARVEGLVAPSPARLAPGPAKDDTASHERA 510
 DB 625 V-----PPGTPKAKSGVKRTIDPATPVYKVP--EAPKSEAEAPVPKPSQDLSRSPQS 676
 QY 511 L-----HDDLESLSENTEFDGILQMAIQSM-----ARPAAPPS 545
 DB 677 LSDTGYSSDGVSSQSEIT--GVVQOEVEQLDSAGVTGPRPSPS 719

RESULT 3

T34239

hypothetical protein F26F12.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34239

R:Willson, R.; Bentley, D.; Gatlung, S.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid F26F12.

A:Reference number: 221493

A:Accession: T34239

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1829 <MIL>

A:Cross-references: EMBL:U55373; PID:NAC25894.1; GSPDB:GN00023; CESP:F26F12.7

A:Experimental source: strain Bristol N2; clone F26F12

C:Genetics:

A:Gene: CESP:F26F12.7

A:Map position: 5

A:Introns: 110/3; 441/3; 801/2; 1244/3; 1693/2; 1784/1

Query Match

7.6%; Score 219.5; DB 2; Length 1829; Best Local Similarity 18.9%; Pred. No. 3.6e-05;

Matches 95; Conservative 56; Mismatches 154; Indels 197; Gaps 15;

QY 22 VDSAFPLHLADHDVYPRDKFOETLHKKEGCPQAFHALLSWLTD-----DSTAILD 76
 DB 85 VCAIIGLTDVEFDYD---EEFFQISNLK-----TFSSITKQILLEANPGTNVSKMYP 134
 QY 77 FWRVLFKDYMLERYGRLPLDPSFKDVLDSQPKKGRPRAPVPRKALVPRPLPTKRKASE 136
 DB 135 MEQVTKREYODHMAAQKPV-----OKQARGSKTPAVSPVPIPPRSAPTKTRGAR 184
 QY 137 EARAAPALTPRGTA SPGSQSLAKAPKPKPSSAEQORLPLGNGIQTMSASVQRAVAMSS 196
 DB 185 RKRDSDA-----PDSDE----- 198
 QY 197 GDVPGARGAVEGLIQVFEESGSKCIQVGEFYTPSKEDS--GSGKNKARSSGKPL 255
 DB 199 -----FEAFIKQD-----EQLEDDLVRKEDAKIRARE 229
 QY 256 VRAKGAOGAPRGGEARLGGQSVAPRLALPSDPQLHOKNEDECAVCRRDGGELICDQGP 315
 DB 230 EKKKGLAARAKAKAKL--EKGE-----EAENNDVCECKQOGELICDQTCR 275
 QY 316 RAFHACLSPRLRIPGTRCSCCLQATVQEVQRAEPRPQEVETPLRPGLSAGE 375
 DB 276 RAVHTVCIDENMEDEPREGWSCHACIEHNGEYVK---EEPAKONDE----- 319
 QY 376 EYRGPRGEPGLAGMDTTLVYKHLRPPRSAPRLPGLDSSALHRLPLCYVREGQONLAPARGC 435
 DB 320 ----- 321
 QY 436 VCGDGTVDYLRCTHCAAFHWRCHPAGTSRPTGT--RCRSCSGDYTPARVEGLA----- 489
 DB 322 ICKETENLLDSCSCSFPAVYCTIDPLTVPRKEETWSCRCETVKRKHKILKICWRMKE 381
 QY 490 -PSPARLAPG--PAKDDTASHER 509
 DB 382 IPYRPLEAGKRASSDDAMLKP 403

RESULT 4

B49515

phosphoprotein 75 - human

C:Species: Homo sapiens (man)

C:Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 10-Sep-1997

C:Accession: B49515

R:Kaderelt, S.; Gewert, D.R.; Galabru, J.; Hovanesian, A.G.; Meurs, E.F.

J. Biol. Chem. 268, 24432-24441, 1993

A:Title: Molecular cloning of two new interferon-induced, highly related nuclear phos

A:Reference number: A49515; MUID:94043285; PMID:7693701

A:Accession: B49515

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-371 <KAD>

A:Cross-references: GB:L22343; NID:g402147; PID:g402148

C:Keywords: phosphoprotein

Query Match 7.5%; Score 219; DB 2; Length 371; Best Local Similarity 25.1%; Pred. No. 7.6e-06; Matches 68; Conservative 37; Mismatches 110; Indels 56; Gaps 9;

```

QY 89 RYGRLOPLDSEPRKVDLSQPRKGRPAVPKALVPPRLP-----TKRKA 134
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104 RKSSEELIDG---TSENNEGRSQKTPSTPRRVGGAASPGHGIOEKLQVNDKVTQRKD 160
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 -----SEERAAAPALTPRGTAASPSQSLKAKPKKPSSAEQORLPLGNGIQTMASAVQR 190
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 161 DSTNMEVEMAVQKARTKCARSS-----RSKEKKKKEDICSSSKRRQKNIHRGKTRKD 215
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 191 AVAMSSGDVPGARGAVEGILLQVEESGSKKCIQ-VGGEFFYTPSKFEESGSKNKARSS 249
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 216 TVDFHCSKLPYTCGEAKGILYKKMKHSSVKCIRNEDGTWLPNPEFEVSGKGRN---A 271
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 250 SGPPPLVAKAGAGAAAGGGEARLQGGGVPAFLALPSPQLHOKNEDECAVCDGGEGLI 309
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 272 KMKRNKICEG-----MTLGE-----LTKRNKNDCEVCCGGQLT 307
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 310 CCDGCPRAFLACLSPPLREITSGTWRCSSC 340
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 308 CCGTCPRVFEHEDCHIRPV-EAKRMMLSCITFC 337
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 5

```

S55259
C:Species: Mus musculus (house mouse)
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jun-2000
C:Accession: S55259
R:Le Douarin, B.; Zechel, C.; Garnier, J.M.; Lutz, Y.; Tora, L.; Pierat, B.; Heery, D.;
EMBO J. 14, 2020-2033, 1995
A:Title: The N-terminal part of TfR1, a putative mediator of the ligand-dependent active
A:Reference number: S55259; M0ID:95262642; PMID:7744009
A:Accession: S55259
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1051 <LEAD>
A:Cross-references: GB:S78221; NID:g998814; PIDN:AAB34290.1; PID:g998815
C:Superfamily: RING finger bromodomain proteins; bromodomain homology
F:48-136/Domain: RING finger homology <RRN>
F:925-981/Domain: bromodomain homology <BR01>

```

Query Match 7.5%; Score 217; DB 2; Length 1051;

Best Local Similarity 27.2%; Pred. No. 2.9e-05; Matches 85; Conservative 46; Mismatches 122; Indels 60; Gaps 15;

```

QY 71 STAILDFWVLFKDYNLERYGRLQPLDSEPRKVDLSQ---PRKGRPAVPKALVPPRLP 126
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 606 SSPMIDLAPVGGSYN-----LPSLP-DIDCSSTIMLDIARKDGVDA---QP 651
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 127 RLPTKRKASEEARAAAPALTPR-GTASPGSQLKAKPP-KKPSSAEQORLPLGNGIQTM 184
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 652 RPPSNRTVO-----SPNSSVSPGLAGPVTMTSVHPPIRSPSASVSGRSSG----- 699
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 185 SASVGRVAVAMSSGDVPGARGAVEGILLQVEESGSKK-----CIQVGEFFYTPSKFEEDS 239
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 700 -SSSPAGADSTHKVPV--MEPIRIKO--ENSGPENYDFPVIVYKQESDEESRPONT 754
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 240 GSGK-----NKARSSSGPKPLVR--AKGAAGAAPG-----GGEARLQGGGVPAFL 282
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 755 NTPRRIITSLILSSSSSEETVLAISDAPDSTGDPGLHDSNNGSGF-WSDASQKSP 813
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 LALPSDPQLHOKNEDECAVCDGGEGLICDGCPRAFHACLSPPLREITSGTWRCSSCGLQ 342
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 814 VAVGETRKEDDNEDECAVCDGGEGLICCEKCPKVFHLTCHVPLTNPSSGEMICTFCRD 873
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 343 ATVOEVQRAEER 355
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 874 LKPEVDYCDVP 886
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 6

T20160
hypothetical protein T14G8.1 - Caenorhabditis elegans

```

C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T20160; T24924
R:McMurry, A.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19231
A:Accession: T20160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1787 <W11>
A:Cross-references: EMBL:267881; PIDN:CAA91798.1; GSPDB:GN00028; CESP:T14G8.1
A:Experimental source: clone C52G5
R:Matthews, P.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19955
A:Accession: T24924
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1787 <W12>
A:Cross-references: EMBL:267884; PIDN:CAA91810.1; GSPDB:GN00028; CESP:T14G8.1
A:Experimental source: clone T14G8
C:Genetics:
A:Gene: CESP:T14G8.1
A:Map position: X
A:Introns: 112/3; 453/3; 597/3; 815/2; 1258/3; 1682/2; 1709/3; 1764/1

```

Query Match 7.1%; Score 206; DB 2; Length 1787;

Best Local Similarity 20.4%; Pred. No. 0.00022; Matches 105; Conservative 58; Mismatches 147; Indels 204; Gaps 22;

```

QY 17 EIVAVDASAPFLHAIADHDV-VPEDKRQETLHLKEKRCQPAFALLSWLTLT-----D 70
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 86 EISAAL-----GLTDVDVDYEOEFQSTTNLKN-----ESSLVKPYLLKVNPGIN 130
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 71 STAILDFWVLFKDYNLERYGRLQPLDSEPRKVDLSQPRKGRPAVPKALVPPRLP 130
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 VTKMPLFYQYKKEF-----QDHMTAHKRSI-----QKQRRKFPV---VPVVPVTPQ 174
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 131 KRKASEBARAAAPALTPRGTAASPSQSLKAKPKKPSSAEQORLPLGNGIQTMASAVQR 190
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 EKTI-----PKTRSSARRK----- 190
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 191 AVAMSSGDVPGARGAVEGILLQVEESGSKKCIQVGEFFYTPSKFEEDS-SGKNKARSS 249
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 -----RDGSG-----EGGHSIDQFEALIKQHEQOQDAEKKEERARIN 231
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 250 SGPPPLVAKAGAGAAAGGGEARLQGGGVPAFLALPSPDQLHOKNEDECAVCDGGEGLI 309
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 232 RAAAKVDKRRRAALESASAKRAR-KEGGV-----EENHQRN---CEVCNODGELM 278
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 310 CCDGCPRAFLACLSPPLREITSGTWRCSSCGLQATVQEVQRAEPRQEPVETPLPG 369
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 279 LCDCTBRAYHVAICDENMEQPEEGDWSG-----PACSEHGPDVLIVEE----- 321
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 370 LRSAGEVVRPGPRL-AGMFTLVYKHLPLAPPSAALPLGLDSSALHPLLCVGPGEQQL 428
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 322 -----EPAAAND----- 329
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 429 APGARCGVCGGTGTVLNRCTHAAAFHWRCHPAGTSRPTGLRCSSGDTVPAP---VE 485
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 330 -----YCRICKETSIWILLCDTCTPSSHYACIDPPLTEIPBEGEMSCRC---IIPERQRIE 382
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 486 GVLA-----PSPARLAPG--PAKDDTASHEP 509
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 KILSMRWKEISYRPLECKEAGEEASKDDVFLKPP 416
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 7

T03455
ALR protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C:Accession: T03455

R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano, Oncogene 15, 549-560, 1997
 A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
 A:Reference number: Z14954; MUID:97388474; PMID:9247308
 A:Accession: T03455
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-4957 <PRA>
 A:Cross-references: EMBL:AF010404; NID:92358286; PIDN:AAC51735.1; PID:92358287
 A:Gene: ALR
 A:Map position: 12
 C:Superfamily: human ALR protein
 C:Keywords: alternative splicing

Query Match 6.7%; Score 194.5; DB 2; Length 4957;
 Best Local Similarity 19.9%; Pred. No. 0.0029;
 Matches 117; Conservative 50; Mismatches 173; Indels 247; Gaps 27;

QY 95 PILDSF-----PKD-VDL--SOPRKGRPPAVPKALVPPRLPTKRKASEARAAAP 143
 DB 530 PALDDFSLGSDTAPLDIDAGSQPEPQPTGSLASELKGSPVL---LDPEELAPVTP 585
 QY 144 AALRP--RGTAAPGSQLAKAPPKPESSAEQRLPLGNGIQTMAS--VQRAVAMSSGDV 199
 DB 586 MEVYPECKQTAGRGSS-----PCEQEPRAPVAPTPPTLLIKSDIYNELISNLSGDA 636
 QY 200 PGAGAVEGILIQVFEESGSKKCIQV--GEFTYPSKFE-----DS 239
 DB 637 SASPFGSEPLLGSPDPPEGGSLS--MELGVSTDVSPARDEGSLRLCTDSLPTDSSLCA 695
 QY 240 GSG-----KNKARSSGPKPLVRAKGA-----OGAPRG--GEARLQGGSV 279
 DB 696 GTATSGKAEGERRRSPARSRIKQGRSSSPGRRRRPGAHGGRGARLAKSTASS 755
 QY 280 PAPLA---LPSDQLHKNED----- 297
 DB 756 IETLVVADIDSSPSKEEEDDDTMTNTVLFSTDKFVLMQMCVCGSFGRAEGHLL 815
 QY 298 -----ECAYCR-----DGELICCGCCRAPIHLA 321
 DB 816 ACSQSGCYHPYCVNSKITVMLLKGRVCYECIYCEVCGQASDPSRLLLCDDCISYHTY 875
 QY 322 CLSPPLREIPSGTWKRS---SCLOATVQEVOPRAEPPRPOEPVETPLPGLRSAGEVR 378
 DB 876 CLDPPLLTVPKGGKCMKVCSCMGCAAS-----PGFHCEMQN-- 913
 QY 379 GPRGPELAGMDTLVYKHLRPARPSAARPLGLDSALHPLLCVPGEGQONTAPGACVCG 438
 DB 914 -----SYTHGCPCASLYTGP-----ICHAPYVEEDL----- 939
 QY 439 DGTDLVLRCTHCAAAFFHRC-----HPAGTSRPGTGLRSCSGDVTAPAVEGYL 488
 DB 940 ----LIQCRHCERMHMAAGCSLFTEDVDH-----APDEGFCVCSQ---PVVAKPVA 985
 QY 489 ASPPARLAGPAKADTASHEPALHRDLESLSLSEHTFDGILQMAIOS 535
 DB 986 PVAPPPELVPMKVEP---EPQYFR-----FEGV--WLTER 1015

RESULT 8
 T03454
 ALR protein - human
 C:Species: Homo sapiens (man)
 C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
 C:Accession: T03454
 R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano, Oncogene 15, 549-560, 1997
 A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolog
 A:Reference number: Z14954; MUID:97388474; PMID:9247308
 A:Accession: T03454
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA

A:Residues: 1-5262 <PRA>
 A:Cross-references: EMBL:AF010403; NID:92358284; PIDN:AAC51734.1; PID:92358285
 C:Gene: ALR
 A:Map position: 12
 C:Superfamily: human ALR protein
 C:Keywords: alternative splicing

Query Match 6.7%; Score 194.5; DB 2; Length 5262;
 Best Local Similarity 19.9%; Pred. No. 0.0031;
 Matches 117; Conservative 50; Mismatches 173; Indels 247; Gaps 27;

QY 95 PILDSF-----PKD-VDL--SOPRKGRPPAVPKALVPPRLPTKRKASEARAAAP 143
 DB 835 PALDDFSLGSDTAPLDIDAGSQPEPQPTGSLASELKGSPVL---LDPEELAPVTP 890
 QY 144 AALRP--RGTAAPGSQLAKAPPKPESSAEQRLPLGNGIQTMAS--VQRAVAMSSGDV 199
 DB 891 MEVYPECKQTAGRGSS-----PCEQEPRAPVAPTPPTLLIKSDIYNELISNLSGDA 941
 QY 200 PGAGAVEGILIQVFEESGSKKCIQV--GEFTYPSKFE-----DS 239
 DB 942 SASPFGSEPLLGSPDPPEGGSLS--MELGVSTDVSPARDEGSLRLCTDSLPTDSSLCA 1000
 QY 240 GSG-----KNKARSSGPKPLVRAKGA-----OGAPRG--GEARLQGGSV 279
 DB 1001 GTATSGKAEGERRRSPARSRIKQGRSSSPGRRRRPGAHGGRGARLAKSTASS 1060
 QY 280 PAPLA---LPSDQLHKNED----- 297
 DB 1061 IETLVVADIDSSPSKEEEDDDTMTNTVLFSTDKFVLMQMCVCGSFGRAEGHLL 1120
 QY 298 -----ECAYCR-----DGELICCGCCRAPIHLA 321
 DB 1121 ACSQSGCYHPYCVNSKITVMLLKGRVCYECIYCEVCGQASDPSRLLLCDDCISYHTY 1180
 QY 322 CLSPPLREIPSGTWKRS---SCLOATVQEVOPRAEPPRPOEPVETPLPGLRSAGEVR 378
 DB 1181 CLDPPLLTVPKGGKCMKVCSCMGCAAS-----PGFHCEMQN-- 1218
 QY 379 GPRGPELAGMDTLVYKHLRPARPSAARPLGLDSALHPLLCVPGEGQONTAPGACVCG 438
 DB 1219 -----SYTHGCPCASLYTGP-----ICHAPYVEEDL----- 1244
 QY 439 DGTDLVLRCTHCAAAFFHRC-----HPAGTSRPGTGLRSCSGDVTAPAVEGYL 488
 DB 1245 ----LIQCRHCERMHMAAGCSLFTEDVDH-----APDEGFCVCSQ---PVVAKPVA 1290
 QY 489 ASPPARLAGPAKADTASHEPALHRDLESLSLSEHTFDGILQMAIOS 535
 DB 1291 PVAPPPELVPMKVEP---EPQYFR-----FEGV--WLTER 1320

RESULT 9
 S58151
 hypothetical protein SPAC2F7.07c - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 21-Jan-2000
 C:Accession: T38555; S58151
 R:Gentles, S.; Churcher, C.M.; Barrell, B.G.; Ralajandram, M.A.; Walsh, S.V. submitted to the EMBL Data Library, July 1995
 A:Reference number: Z21799
 A:Accession: T38555
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-607 <GE2>
 A:Cross-references: EMBL:Z50142; NID:91052783; PIDN:CAA90494.1; PID:91052790; GSPDB:G
 A:Experimental source: strain 972h-; cosmid c2F7
 C:Gene: SPAC2F7.07c
 A:Map position: 1

Query Match 6.4%; Score 186.5; DB 2; Length 607;

Best local similarity 22.7%; Pred. No. 0.001;
Matches 83; Conservative 44; Mismatches 132; Indels 107; Gaps 18;

QY 151 TASPSSQAKAPKPKRESSAEQOQLPLNGIOTMSASVORAVAMSSGDVGARGAVEGIL 210
DB 169 TAKDLSDISSSMKKANSSK-----PLFSGKLTKFNT-----PPTSEVTENNYY 215
QY 211 IQQFESSGSKKCIQVGGFFYPPSPFEDSGSKNNARSSSGPKPLVANKAQAAGAGGGE 270
DB 216 TRNVTVYSNOK---HLGNE---SENFNDM---EGRAEDISSNELL----- 251
QY 271 ARIQGGGVAPLALPSPDQHLQKNEDECANCRDGEILLCCDCCPRAPHLACLSPURE- 329
DB 252 -----PPEEYR-----TRNNDYCSACHGPNFLCCTCETCPNSFHTCTIDPIEEK 297
QY 330 -IPSGTMRSCSLQATV-----OEVPRAEPR-----POEPPVEPRLP 368
DB 298 NLPDQAWYCNCKHHSILVNEDEGELESNVKEEGTMVVMQLCTYISHNFIQHLPH 357
QY 369 GL-----RSAGEVVRGPPPEPLAGMDTLYYKHLPAF-----PSADPLPGLDSSALHPLLC 419
DB 358 SISSEFFRGVSGVMGEYIE-----TDVLKHLKSSRRSNGEERDPLL-LKSSGTPIIC 409
QY 420 VGPEQQQLLARGARGVCG-DGTDVLRCTHCAAFHMRCHFRPAGTSRPTGLR---CRSC 475
DB 410 F-----RCHKSAVYSQSLIACDYCNMYHDPDLNPLATLP-SNLKMKCPNH 456
QY 476 SGDYTP 481
DB 457 SDHYTP 462

RESULT 10

562475
PND finger protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jan-2000
C:Accession: T37778; 562475
R:Bedcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z21745
A:Accession: T37778
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-404 <BA2>
A:Cross-references: EMBL:Z54366; NID:91019812; PIDN:CAA91193.1; PID:91019817; GSPDB:GN00
A:Experimental source: strain 972h-; cosmid c16c9
A:Genetics:
A:Gene: SPAC16C9.05
A:Map position: 1L

Query Match 6.2%; Score 179.5; DB 2; Length 404;
Best local similarity 25.5%; Pred. No. 0.0017;
Matches 67; Conservative 27; Mismatches 94; Indels 75; Gaps 12;

QY 294 KNEDECAVCRDGGELICDCCPRAPHLACLSPPL-REIPSGTMRSCSLQATVOEVQPR 351
DB 115 RNVDYCSACGGRLGFLICCEGPCSFHLSCLEPRLPENIPESWFCVYC---SISNHP- 170
QY 352 AEPRRQEP-----PVETPLRPLGRSAGEEV-RGPPG---EPLAGMDTTL 392
DB 171 -----PKHPLSTWSQLYDWIDSONPSOYRLPDDLVHYFHGISRGDTGAVKETEGEDMT- 223
QY 393 VYKHLPAAPSAAPRLGDDSSALHPLLCVPRGQQLNARGARGVCGDGDID---VLKCH 448
DB 224 --DEFSALPTGSSITNL-----ATCGYCSKPSMGACWVYGCOL 259
QY 449 CAAAFHMRCHFRPAGTSRPTGLRCRSCSGDYTPAPVEGVLAAPPAALGAPAKDTASH- 507
DB 260 CDFEYHKCKEHA-----KKCSHD--SIGKKGMKRVKKNVIVTRPLVLDITSTNT 306
QY 508 -EPALHRDDLESLSEHTFGCIL 529

DB 307 LNPRVMTSGWQFLMGEPFSDLL 329

RESULT 11

T42730
Bassoon protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42730
R:Dieck, S.; Sammarti-Vila, L.; Langreese, K.; Richter, K.; Kindler, S.; Soyke, A.; W.
J. Cell Biol. 142, 499-509, 1998
A:Title: Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively localized
A:Reference number: Z22249; MUID:98345363; PMID:9679147
A:Accession: T42730
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3942 <DIE>
A:Cross-references: EMBL:Y17034; NID:93413809; PIDN:CAA76598.1; PID:93413810
A:Experimental source: strain 129 SVJ
C:Genetics:
A:Map position: 9P1
A:Introns: 72/2; 208/3; 505/3; 675/3; 2889/3; 3582/1; 3851/3; 3886/1; 3930/1
A:Note: bassoon
C:Function:
A:Description: may be involved in cytomatrix organization at the site of neurotransmi
A:Note: component of the presynaptic cytoskeleton
C:Keywords: coiled coil; zinc finger

Query Match 6.2%; Score 178.5; DB 2; Length 3942;
Best local similarity 22.1%; Pred. No. 0.02;
Matches 118; Conservative 50; Mismatches 230; Indels 135; Gaps 24;

QY 101 PKDYDLSOPKRGKRPVAVKALVPPRLPTRKKAASEEAAAAPALTRGTASPG-SQLK 159
DB 250 PAKQPLKPRDERSPRG-PGATQSGPR-----QEAARATSVSPPT-QATPAPEVGRVS 301
QY 160 AKPPKPPRESSAEQOQLPLNGIOTMSASVORAVAMSSGDVGARGAVGILLQVFEESG 219
DB 302 PQPLSLTKRPTAEERPRAG-----EAQGSATTVSGLAGAGT-----QBL 344
QY 220 SKKCIQVGEFTY-----PSK-----FEDSG--SGKNKARSSGP 252
DB 345 TGRKLFGLGASLLTQASTLMSVQPADTQGGPSKPGPKYIFSDASKRAGPRPESGPGP 404
QY 253 KPLVRAAGAGA--APGGEARLQOQGSVPR-----IALPSDPLQKNEDEC 299
DB 405 GPPPGAKTEPGARTGPGSGPGLAKTGCTASPKGRHAQASAKMAKPKTMPKERAAC 464
QY 300 AVC-----RDGCELICDCCP-RAVHFLACLSPPLREIPSGTMRSCS-----LQATV 345
DB 465 PLCOAEINVSQRGANNTCTACKLQVCNLCGFNPTPLVETKTEWLCINCTKRLLESL 524
QY 346 QEVQRAEERPRQEPVETPLRPLGRSAGE---EVRGPPG-----EPLAGMDT 390
DB 525 GEPAP-LPLPTPOQPPAGVP---HRAAGAPRLQKQKGGQIGQPSGLSPAKASPAQTATAS 579
QY 391 TLVYKHLPAAPSAAPRLGDDSSALHPLLCVPRGQQLNARGARGVCGDGDVLRCTHCA 450
DB 580 PQATKASPAQTATASPAQTATASPAQTATASPAQTATASPAQTATASPAQTATASPAQTATAS 636
QY 451 AAFHMRCHFRPAGTSRPTGLRCRSCSGDYTPAPVEGVLAAPSP-----ARLAPRAKD 502
DB 637 -----PETVVPETP---KAKSGVKRTDPAVPAVKPVEAPKGEAEPEVPKPYSO 684
QY 503 DTASHEPAL-----HRDDESLSEHTFGDILQALQISM-----ARPAAPPS 545
DB 685 DLSRSPQSLSDTGYSSDGVSSQSEIR--GVYQGEVYQDLSAGVGTGRPREPS 735
RESULT 12
T02518
hypothetical protein At2g37520 [imported] - Arabidopsis thaliana
N/Alternate names: hypothetical protein F13M22.2

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02518; G84793
R:Rounsley, S.D.; Kaul, S.; Liu, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Sheph, M.; Vanden, S.E.; Umayam, L.; Tallon, L.; Nause, D.; Nleman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: T02518
A:Reference number: 214677
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-825 <R0U>
A:Cross-references: EMBL:AC004684; NID:g3236234; PID:g3236235
A:Experimental source: Cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Sheph, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Sheph, M.; Vanden, S.E.; Umayam, L.; Tallon, L.; Nause, D.; Nleman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: T02518
A:Reference number: 214677
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-825 <SNO>
A:Cross-references: GB:AE002093; NID:g3236235; PID:AMC23623.1; GSPDB:GND0139
C:Genetics:
A:Gene: F13M22.2; A12937520
A:Map position: 2
A:Intons: 189/3; 212/3; 282/3; 304/1; 366/2; 388/2; 415/3; 433/2; 489/1; 550/2; 569/3;
C:Superfamily: Arabidopsis thaliana hypothetical protein F4P12.380

Query Match
Best Local Similarity 5.9%; Score 171; DB 2; Length 825;
Matches 29; Conservative 18; Mismatches 36; Indels 2; Gaps 1;

Y 295 NEDECAVCRGGEICDCCPRRAHACLSPPLEIRSGTSCSCLOATVOEQPRAEE 354
Db 463 SDMCSTCGGGDLICAGCPQAHTACLK--FQSMPEGTWCSCNDGPISKKATTTD 520
Y 355 PRQEPPEVETPLPGLRSAGEEVG 379
Db 521 PEGNARPIVRLSRVYKAPESDIOG 545

RESULT 13
QOBE3
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
C:Accession: A03742
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A:Reference number: A93065; MUID:85035713; PMID:6092825
A:Accession: A03742
A:Molecule type: DNA
A:Residues: 1-660 <BAN>
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-211, 1984
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A:Reference number: A03794; MUID:84270667; PMID:6087149
A:Contents: annotation; protein coding region
C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52
C:Superfamily: human herpesvirus 4 BHLF1 protein

Query Match
Best Local Similarity 5.8%; Score 168.5; DB 1; Length 660;
Matches 120; Conservative 26; Mismatches 192; Indels 175; Gaps 25;

Y 109 PRKGRP-----PAVPAALVPPLPTKRKASEEAAAPALPR-----GTASPGS 156
Db 172 PGAGRGPGPTGGRPAAPGA-----PCTPA--APGCGGAAPVSGATPHPERGSGPADPPA 225
Y 157 QLKAKP-----PKPPSSAEQORLPLGNGIQTMSASVQRAVAMSSGDVPGAR----- 203

Db 226 AARLPPEQERLPODLAAACRCPAPPTTSGAAQRTNRRPCCPSARPPCPTWR 285
Y 204 ---GAVEGILLIOVFESGSKKCIQVGGFEFTPSKFEEDSGGKNKARSSGKPLVRAG 260
Db 286 RRSQAQKG-----HPP-----PGAGQ-RPSPTGGRP--AAPG 315
Y 261 AQC--AARGGEARLGQGSVPAPL--ALPSDPLQKNEDECAVCRDGGELICCGGPR 316
Db 316 APGTPAAPCGGGAAPVSGATPHPERGSGPADPPAARLPREQERLPODLAAACRCPA 375
Y 317 APHLACSLPLR-----ETPSG-----TWCSCLQATVQEVQRAEER 355
Db 376 G-----PPPTSCGAQAORTHRRPPGCRSARNDCPTTHRRSGAQ----- 416
Y 356 RQEPPEVETPLPGLRSAGEEVGRPREPLAGMDTLVYKILPAPPSAARPGLDSSALH 415
Db 417 RGNRPQAGRPSGPTGGRPAARPGATPRA-----PQGGGAAPV--SGATPH 463
Y 416 PLLCVG-----PEGO-----ONLAPGARC-----GVCGDGTD----- 442
Db 464 PERGSGPADPPAARLPPEQERLPODLAAACRCPAPPTTSGAAQRTNRRPCCPR 523
Y 443 VLCTHCAAFHMRCHRPACTSRPGTGLRCRSCG-----DVTAPVEGLAPS 491
Db 524 SANPCCPRTWRRRSQAQRHPRPGAGORPSGPTGGRPAARPGATPRAARPGGGAAPV 583
Y 492 PARLAP-----GPAKDDTAS-----HEPALHRD 514
Db 584 GATPHPERGSGPADPPAARLPPEQERLPOD 616

RESULT 14
T45070
protein kinase homolog R31240.1 [imported] - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T45070
R:Lamerdin, J.; McCreedy, P.; Stillwagen, S.; Ramirez, M.; Carrano, A.
submitted to the EMBL Data Library, November 1996
A:Description: Characterization by genomic sequence analysis of a gene-rich 111 kb re
A:Reference number: 222906
A:Accession: T45070
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1237 <LAW>
A:Cross-references: EMBL:AD000092; PID:AA51171.1
A:Experimental source: cell line 5HL2-B; fibroblast
C:Genetics:
A:Map position: 19p13.2
A:Note: Intron positions not resolved (incomplete sequence)

Query Match
Best Local Similarity 5.8%; Score 168; DB 2; Length 1237;
Matches 111; Conservative 41; Mismatches 180; Indels 88; Gaps 22;

Y 116 PAVPAALVP-----PRLPTKRKASEEAAAPALPRGTASPSGQTKAKPKPPRESS 169
Db 870 PLSPPLAHPPTQASPPPLPGHTVGSSTHTGSPFAKLHSSPPVRRPRASAEPRSPPLK 929
Y 170 AEOGRPLGNGIQTMSASVQRAVAMSSGDV--PGARGAVEG--TLIQVFESGSKKCIQV 226
Db 930 RVQSAEKLG---ASLSADKKGALRHSLEVGHPRKDFHGEALHLSLAESDGEPRVVG 986
Y 227 GGEFTPSKFEEDSGGKNKARSSGKPLVRKGA-----QGAAPCGGGAARLGQGS 278
Db 987 LG---APQVAVNRKRGQESPLSGADPLL--PEGASRPVSKKEESPAGACATPPRAT 1042
Y 279 VPAPLAPSD--OLQKNEDECAVCRDGGELICDCCPRAPHLACLSPLEIRPSGTR 336
Db 1043 TPGRTLELDVGCGRHGSVQRE-----DG-----TGMARAVAKAALS-PVQENETG--R 1089
Y 337 CSSCLQATVQEVQRAEERPRQEPPEVETPLPGLRSAGEEVGR--PGEPLA--GMD--- 389

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:45:58 ; Search time 32 Seconds
(without alignments)
3509.240 Million cell updates/sec

Title: us-09-509-595b-2

Perfect score: 2902 1 MATDALARLRLLRLHREIAY.....DGLIOWAQSMARPAAPPSS 545

Sequence: 1 MATDALARLRLLRLHREIAY.....DGLIOWAQSMARPAAPPSS 545

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeoprotein:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2687	92.6	515	4 075745	075745 homo sapien
2	2005	69.1	552	11 0920E3	0920E3 mus musculu
3	1991	68.6	548	11 0921W9	0921W9 mus musculu
4	1989.5	68.6	551	11 0921X0	0921X0 mus musculu
5	1975.5	68.1	547	11 0921W8	0921W8 mus musculu
6	1853.5	63.9	493	11 0921W7	0921W7 mus musculu
7	1839.5	63.4	489	11 0921W5	0921W5 mus musculu
8	1838	63.3	492	11 0921W6	0921W6 mus musculu
9	1824	62.9	488	11 0921W4	0921W4 mus musculu
10	1438.5	49.6	409	11 0921W3	0921W3 mus musculu
11	1424.5	49.1	405	11 0921W1	0921W1 mus musculu
12	1423	49.0	408	11 0921W2	0921W2 mus musculu
13	1409	48.6	404	11 0921W0	0921W0 mus musculu
14	270.5	9.3	2000	4 0921A10	0921A10 homo sapien
15	253	8.7	1954	4 0821D10	0821D10 homo sapien
16	229	7.9	689	4 0921B58	0921B58 homo sapien

17	221.5	7.6	3938	11 088778	088778 ratu norv
18	219.5	7.6	1829	5 019815	019815 caenorhabdi
19	219	7.5	408	4 014977	014977 homo sapien
20	217.5	7.5	1214	4 09B004	09B004 homo sapien
21	217	7.5	516	11 088154	088154 mus musculu
22	214	7.4	429	11 09D108	09D108 mus musculu
23	214	7.4	1209	11 0924W6	0924W6 mus musculu
24	210.5	7.3	245	4 09H930	09H930 homo sapien
25	209	7.2	885	4 096795	096795 homo sapien
26	197.5	6.8	607	10 091G55	091G55 oryza sativ
27	196	6.8	5038	11 0901X7	0901X7 mus musculu
28	194.5	6.7	4957	4 014687	014687 homo sapien
29	194.5	6.7	5262	4 014686	014686 homo sapien
30	192	6.6	680	4 0968B5	0968B5 homo sapien
31	188	6.5	5085	11 09JX86	09JX86 ratu norv
32	187.5	6.5	888	5 09M410	09M410 drosophila
33	186.5	6.4	1139	5 09VDS5	09VDS5 drosophila
34	186	6.4	704	4 0960T6	0960T6 homo sapien
35	184.5	6.4	878	5 08T001	08T001 drosophila
36	184.5	6.4	3851	4 043161	043161 homo sapien
37	184.5	6.4	3926	4 09UPA5	09UPA5 homo sapien
38	184	6.3	2228	10 046579	046579 arabidopsis
39	182.5	6.3	356	11 08VECB	08VECB mus musculu
40	182.5	6.3	810	10 0945C8	0945C8 zea mays (m
41	182	6.3	5147	4 09Y6V0	09Y6V0 homo sapien
42	181	6.2	627	5 08SYJ8	08SYJ8 drosophila
43	181	6.2	887	5 016102	016102 drosophila
44	180	6.2	1133	5 09U377	09U377 drosophila
45	179.5	6.2	812	10 09MB44	09MB44 oryza sativ

ALIGNMENTS

RESULT 1
ID 075745 PRELIMINARY; PRT; 515 AA.
AC 075745;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE AIRE protein.
GN AIRE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98061087; PubMed-9398840;
RA Aaltonen J., Björns P., Perheentupa J., Horelli-Kuitunen N.,
RA Palot A., Peltonen L., Lee Y.S., Francis F., Hennig S., Thiel C.,
RA Lehrach H., Yaspo M.L.;
RT "An autoimmune disease, APECED, caused by mutations in a novel gene
RT featuring two PHD-type zinc finger domains.";
RL Nat. Genet. 17:399-403(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Lee Y.S., Francis F., Hennig S., Thiel C., Reinhard R., Lehrach H.,
RA Yaspo M.L.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ009610; CA08759.1; -
DR InterPro; IPR000770; SAND_domain.
DR InterPro; IPR004865; SPI00.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF01342; SAND; 1.
DR Pfam; PF01172; SPI00; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00258; SAND; 1.
SQ SEQUENCE 515 AA; 54600 MW; 2AF5B3DC28605A3A CRC64;

Query Match

92.6%; Score 2687; DB 4; Length 515;

Best Local Similarity 94.5%; Pred. No. 3.8e-174;
Matches 515; Conservative 0; Mismatches 0; Indels 30; Gaps 2;

```

QY 1 MATDALARLRLHRTETAVAVDSAPFLHLALADHDVPEDEKFOETLHLKEKGGCPQAFH 60
DB 1 MATDALARLRLHRTETAVAVDSAPFLHLALADHDVPEDEKFOETLHLKEKGGCPQAFH 60
QY 61 ALLSWLITQDSTALIDFWRLFKDYNLERGRLOPLIDSPKDVDSQPRKGRPPAVPK 120
DB 61 ALLSWLITQDSTALIDFWRLFKDYNLERGRLOPLIDSPKDVDSQPRKGRPPAVPK 120
QY 121 ALVPPRLPTKRKASEEARAAPALTPRGTAPOGSQLAKPKPKESSAEQORLPLGNG 180
DB 121 ALVPPRLPTKRKASEEARAAPALTPRGTAPOGSQLAKPKPKESSAEQORLPLGNG 180
QY 181 IOTMSAVQRAVAMSSGGVPGARGAVEGILLIOVFESGSKKCIQVGEFYTPSKFEDSG 240
DB 181 IOTMSAVQRAVAMSSGGVPGARGAVEGILLIOVFESGSKKCIQVGEFYTPSKFEDSG 240
QY 241 SGKKARSSSGPRPLVRKAGAGAAPGGEGARLGGGVPAPLALPSDQLHOKNDECA 300
DB 241 SGKKARSSSGPRPLVRKAGAGAAPGGEGARLGGGVPAPLALPSDQLHOKNDECA 300
QY 301 VCRBGELICDGGCPRAFHACLSPLREIPSGTWRCSSCIQATVOEQVOPRAEPPROP 360
DB 301 VCRBGELICDGGCPRAFHACLSPLREIPSGTWRCSSCIQATVOEQVOPRAEPPROP 360
QY 361 PVETPLPGRSAGEEVRGPEGLAGDITLVYKHLPAAPSADPLGLDSSALHPLLCV 420
DB 361 PVETPLPGRSAGEEVRGPEGLAGDITLVYKHLPAAPSADPLGLDSSALHPLLCV 420
QY 421 GPEGQONLARGARGV-----W-----TGLCRSSGCVT 480
DB 421 GPEGQONLARGARGV-----W-----TGLCRSSGCVT 480
QY 481 PAVEGVLAAPARLAPGPAKDDTASHEPALHRDLESLSHEHFDGILQMAISMAPPA 540
DB 481 PAVEGVLAAPARLAPGPAKDDTASHEPALHRDLESLSHEHFDGILQMAISMAPPA 540
QY 541 APPPS 545
DB 541 APPPS 515

```

RESULT 2

```

QY 0920E3 PRELIMINARY: PRT: 552 AA.
AC 0920E3:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Aire protein (AUTOIMMUNE regulator).
GN Aire.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=129;
RA Katin B., Schweigler M., Wertz K., Poulson R., Christensen H.M.,
RA Rosenthal A., Leirach H., Yaspo M.L.;
RT "The mouse Aire gene: comparative genomic sequencing, gene
RT organization and expression.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=129SV;
RA MEDLINE=9916080; PubMed=10049735;
RA Miltaz L., Rossier C., Heino M., Petersen P., Krohn K.J.E., Gos A.,
RA Morris M.A., Shimizu N., Antonarakis S.E., Scott H.S.;
RT "Isolation and characterization of the mouse Aire gene.";
RL Biochem. Biophys. Res. Commun. 255:483-490(1999).

```

```

RN 13
RP SEQUENCE FROM N.A.
RC STRAIN=B6; TISSUE=THYMUS;
RX MEDLINE=99168902; PubMed=10049587;
RA Wang C.Y., Shi J.D., Davoodi-Semlomi A., She J.X.;
RT "Cloning of Aire, the mouse homologue of the autoimmune regulator
RT (Aire) gene responsible for autoimmune polyglandular syndrome type 1
RT (ASPI).";
RL Genomics 55:322-326(1999).
RN 14
RP SEQUENCE FROM N.A.
RA Halonen M., Peltto-Huikko M., Palvimo J., Bjoerses P., Pelttonen L.,
RA Umanen I., Kolmer M.;
RT "Expression of the mouse Aire.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN 15
RP SEQUENCE FROM N.A.
RA Shi J.D., Wang C.Y., Marron M.P., Ruan Q.Q., Huang Y.Q., Dettler J.C.,
RA She J.X.;
RT "Chromosomal Localization and Complete Genomic Sequence of the Murine
RT Autoimmune Regulator Gene (Aire).";
RL Autoimmunity 0:0-0(1999).
RN 16
RP SEQUENCE FROM N.A.
RC STRAIN=MOD, AND SJL;
RA Shi J.D., Wang C.Y., Marron M.P., Ruan Q.Q., Huang Y.Q., Dettler J.C.,
RA Davoodi-Semlomi A., She J.X.;
RT "Complete genomic sequence, gene structure and localization of the
RT mouse Aire gene." to the EMBL/GenBank/DBJ databases.
RN 17
RP Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ132243; CAB36909.1; -
DR EMBL: AJ007715; CAA07620.1; -
DR EMBL: AF079536; AAD20444.1; -
DR EMBL: AJ243821; CAB66141.1; -
DR EMBL: AF105002; AAD46421.1; -
DR EMBL: AF128773; AAF36482.1; -
DR EMBL: AF128772; AAF36481.1; -
DR MGI: 1338803; Aire.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000770; SAND domain.
DR InterPro: IPR004865; Sp100.
DR InterPro: IPR001865; ZnF-PHD.
DR InterPro: IPR001841; ZnF-ring.
DR Pfam: PF00628; PHD; 2.
DR Pfam: PF01342; SAND; 1.
DR SMART: SM00172; Sp100; 1.
DR SMART: SM00001; EGF-like; 1.
DR SMART: SM00249; PHD; 2.
DR SMART: SM00184; RING; 2.
DR SMART: SM00258; SAND; 1.
SQ SEQUENCE 552 AA; 59042 MW; BF30F66B71239A CRC64;

```

Query Match 69.1%; Score 2005; DB 11; Length 552;

Best Local Similarity 71.8%; Pred. No. 6.3e-128;

Matches 393; Conservative 37; Mismatches 108; Indels 10; Gaps 6;

```

QY 4 DAALRLRLHRTETAVAVDSAPFLHLALADHDVPEDEKFOETLHLKEKGGCPQAFHALL 63
DB 5 DGMLRLRLHRTETAVAVDSAPFLHLALADHDVPEDEKFOETLHLKEKGGCPQAFHALL 64
QY 64 SWLITQDSTALIDFWRLFKDYNLERGRLOPLIDSPKDVDSQPRKGRPPAVPKALY 123
DB 65 SWLITQDSTALIDFWRLFKDYNLERGRLOPLIDSPKDVDSQPRKGRPPAVPKALY 124
QY 124 PPRPLPTKRKASEEARAAPALTPRGTAPOGSQLAKPKPKESSAEQORLPLGNGTGT 183
DB 125 LPPRPLPTKRKASEEARAAPALTPRGTAPOGSQLAKPKPKESSAEQORLPLGNGTGT 184
QY 184 MSASVQRAVAMSSGGVPGARGAVEGILLIOVFESGSKKCIQVGEFYTPSKFEDSGSG 242
DB 185 MASVQRAVAMSSGGVPGARGAVEGILLIOVFESGSKKCIQVGEFYTPSKFEDSGNL 244
QY 243 KKKARSSSGPRPLVRKAGAGAAPGGEGARLGGGVPAPLALPSDQLHOKNDECAVC 302

```

[illegible]

RESULT 3					
ID	Q9JLM9	PRELIMINARY:	PRT:	548 AA.	
AC	Q9JLM9				
DT	01-OCT-2000	(TREMBLrel. 15, Created)			
DT	01-OCT-2000	(TREMBLrel. 15, last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	Autoimmune regulator.				
CN	AIRE.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=B6;				
RA	Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;				
RT	"Expression and alternative splicing of the mouse autoimmune regulator gene (Aire)." ;				
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF128116; AAC36461.1; -.				
DR	MGD; MGI:1338803; Aire.				
DR	InterPro: IPR000345; CYC_heme_bind.				
DR	InterPro: IPR000561; EGF-like.				
DR	InterPro: IPR000770; SAND_domain.				
DR	InterPro: IPR004865; Sp100.				
DR	InterPro: IPR001965; ZnF_PHD.				
DR	InterPro: IPR001841; ZnF_ring.				
DR	Pfam: PF00628; PHD; 2.				
DR	Pfam: PF01342; SAND; 1.				
DR	Pfam: PF03172; SP100; 1.				
DR	SMART: SM00001; EGF_like; 1.				
DR	SMART: SM00249; PHD; 2.				
DR	SMART: SM00184; RING; 2.				
DR	SMART: SM00258; SAND; 1.				
DR	PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.				
SQ	SEQUENCE 548 AA: 58631 MW: 0B052443C5EDCB6 CRC64;				
Query Match 68.6%; Score 1991; DB 11; Length 548;					
Best Local Similarity 71.6%; Pred. No. 5..5e-127;					
Matches 394; Conservative 37; Mismatches 105; Indels 14; Gaps					
OY	4 DAALRLRLRHTETAAVDAVSAPPLIHALLADHDVPEDKFOEFTLHKKEGCPQAFHALL 63				
DsB	5 DGMRLRLRLRHTEETAVIDSAFPPLIHALADHDVDVEDDFOETLRKEKRGKCCPQAFHALL 64				
OY	64 SMLLTQDSSTAIIADPFVRVLEKDYNLERGRGLPTISFFPKDVDSOPRKGRKKPAVPKALV 123				
DsB	65 SWLTRDSGAILDPFRKLTKDYNLSEYSRSLHSITDSFPKDVDDNSQSKKKPLAGPPAAV 124				

QY	124	PPPLPTKRRASEEAPAAAPALPRRGTAISPGSOLAKPRKPESSAEORLPLNGIOT	183
Db	125	LPPLPPKRRALREPRATPPATTLASKSVSSPSHLLTKTPKKEDGNDLESOHLPKMGLOT	184
QY	184	MSASQRAVAMSSSDVDGARGAVEGILLIQOVFEESGSKKCIQVGEHYTPSKFED--SGSG	242
Db	185	MAASQRAVAVASGDVDGTGSAVNGILIQOVFEESGRSKKCIQVGEHYTPKPKFEDSGNLL	244
QY	243	KNKRRSSGPRPLVYRAKGAGCAAPGGEGEARLGGQGSVPAPLALPSPDPLHKKKNEDECAYC	302
Db	245	KNKRRSSSLKPVYRAGAGGR---DEQKVGQCCGVPPLPSLPSEPOVYKKKNEDECAYC	300
QY	303	RDGELLICDCCGAPFHLACISPLREIIPSGTWKSCCLOATVOEOPRAEERPOPEPVY	362
Db	301	HDGELLICDCCGAPFHLACISPLLOGLIPSGLRKSCCLOGRQONLSQEVSRPPELPA	360
QY	363	ETPLPGLRSAGEEVRGPRGEBPLAGMDTTLVYKHLPAAPSAAPLPGLDSSALHPLLCVGP	422
Db	361	ETPLVGLRSASEKTRGPRSELKASSDAATVYVLLHPHRAAPL--LEPSALCPLLSAGN	418
QY	423	EGGQNNLAPGARGCGVGGGTDLRGTHTCAAPFHMHCCHPACTSRPGTGLRSCSGGVTPA	482
Db	419	EGRGPRAPASACSVCGGTEVLRCAHQAALAHMHCCHPPTLAARGTNLLRKCKSSASSTPR	478
QY	483	P-VEGVLAIP--SPARKLGGPAK--DDTASHEPALRDLDLESILSEHTFDGILLQWAIQMSAR	538
Db	479	PGTGGEAVNTPSGRPARPARGLAKVGDSDASHDPVLHHRDLDLESILNHSRFDGILLQWAIQMSR	538
QY	539	PAA---PPPS	545
Db	539	PLAETPPSS	548

RESULT	4
O9JLX0	
ID	O9JLX0 PRELIMINARY; PRT; 551 AA.
AC	O9JLX0;
DT	01-OCT-2000 (TREMBLrel. 15, Created)
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE	Autoimmune regulator.
GN	AIRE.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=B6;
RA	Ruan O.G., Wang C.Y., Shi J.D., She J.X.;
RT	"Expression and alternative splicing of the mouse autoimmune regulator
RL	gene (Aire).";
SR	Submitted (FEBS-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AF128115; AAF36460.1; -
DR	MED; MG1:1338803; Aire.
DR	InterPro: IPR000345; Cytc_heme_bind.
DR	InterPro: IPR000561; EGF-like.
DR	InterPro: IPR00770; SAND_domain.
DR	InterPro: IPR004865; Spi100.
DR	InterPro: IPR001965; Znf_PHD.
DR	InterPro: IPR001841; znf_ring.
DR	Pfam: PF00628; PHD. 2.
DR	Pfam: PF01342; SAND. 1.
DR	Pfam: PF03172; SPI00. 1.
DR	SMART: SM00001; EGF_like. 1.
DR	SMART: SM00249; PHD. 2.
DR	SMART: SM00184; RING. 2.
DR	SMART: SM00258; SAND. 1.
DR	PROSITE: PS00190; CYTOCHROME_C_1.
DR	SEQUENCE 551 AA; 58913 MW; 77c75e773b48b72c CRC64;

Best Local Similarity 71.6%; Pred. No. 7e-127;
Matches 394; Conservative 37; Mismatches 108; Indels 11; Gaps 7;

```

QY 4 DAALRLRLRLHRTETAVAVDSAPFLHLADHDVPEDEKFOETLHLKEEGCPQAFHALL 63
Db 5 DGMRLRLRLHRTETAVAVDSAPFLHLADHDVPEDEKFOETLHLKEEGCPQAFHALL 64
QY 64 SWLLTQDSTALIDFWRVLEFKDYNLERYGRLQPLIDSEPKVDLSQPRKGRPPAVPKALV 123
Db 65 SWLLTRDSGALIDFWRVLEFKDYNLERYGRLQPLIDSEPKVDLSQPRKGRPPAVPKALV 124
QY 124 PPPRLPTKRKASEARAAAPALPRGTASPGSQLAKAPKPKPESSAEQRLPLGNGIOT 183
Db 125 LPPRPPTKRKALEEPRAATPATLASKSVSSPSGSHLTKPRKPPDGNLESOHLPLGNGIOT 184
QY 184 MSASVORAVAMSSGDVPGARAGAVEGILIQOVFESGSKKCIQVGEFYTSPSKFED-SGSG 242
Db 185 MAASVORAVAVASGDVPGTGAVEGILIQOVFESGSKKCIQVGEFYTSPSKFEDPSGNL 244
QY 243 KKKARSSSGPKPLVRAKGAOGAARPGGEARLGQGSVPAPALPSPDQLHOKNDECAVC 302
Db 245 KKKARSSSLKLPVRAKGAOQTITIGRDQKVGQCCVPLPSLSEPOVNO-NEDECAVC 303
QY 303 RDGELICDGCPRAFHLACLSPPLREIPSGTWRCSCLQATVOEVPRAEPRPOEPPV 362
Db 304 HDGELICDGCPRAFHLACLSPPLREIPSGTWRCSCLQGRVQONLSQPEVSRPELPA 363
QY 363 ETPPLPGLRSAGEVRGPRGPELAGMDTTLVYKHLPAAPSAPLPGLDSSALHPLLCVGP 422
Db 364 ETPILVGLRSASEKTRGPRRELKASSDAAYTYVNLPAHPAAPL--LEPSALCPILLSAGN 421
QY 423 EGOONLAPGARCGVCGDGTDLVLRCTHCAAFHWRCHFPAGTSRPGTGRCSSGSDVTPA 482
Db 422 EGRPGAPARSVCYCGDGTDLVLRCAHCAAFHWRCHFPAGTSRPGTGRCSSGSDSTPT 481
QY 483 P-VEGVLP-SPARLAPGPAK--DDTASHEPALHRDLSLSEHTFEDGILQMAIOSMAR 538
Db 482 PGTBEAVPTSGPRAPAGLAKVGDSDASHDPVLRHDDLSLSEHTFEDGILQMAIOSMSR 541
QY 539 PAA---PPPS 545
Db 542 PLAEPTPPSS 551

```

RESULT 5
Q9JLM8 PRELIMINARY: PRT: 547 AA.
AC Q9JLM8:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Autoimmune regulator.
GN AIRE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6.
RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
RT "Expression and alternative splicing of the mouse autoimmune regulator gene (Aire).";
DR EMBL: AF128117; AAF36462.1; -;
DR MGD: MGI:1338603; Aire.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000770; SAND domain.
DR InterPro: IPR004865; SP100.
DR InterPro: IPR001965; Znf_PHD.
DR InterPro: IPR001841; Znf_fing.
DR Pfam: PF00628; PHD; 2.

Query Match 68.1%; Score 1975.5; DR 11; Length 547;
Best Local Similarity 71.5%; Pred. No. 6.2e-126;
Matches 393; Conservative 37; Mismatches 105; Indels 15; Gaps 8;

```

QY 4 DAALRLRLRLHRTETAVAVDSAPFLHLADHDVPEDEKFOETLHLKEEGCPQAFHALL 63
Db 5 DGMRLRLRLHRTETAVAVDSAPFLHLADHDVPEDEKFOETLHLKEEGCPQAFHALL 64
QY 64 SWLLTQDSTALIDFWRVLEFKDYNLERYGRLQPLIDSEPKVDLSQPRKGRPPAVPKALV 123
Db 65 SWLLTRDSGALIDFWRVLEFKDYNLERYGRLQPLIDSEPKVDLSQPRKGRPPAVPKALV 124
QY 124 PPPRLPTKRKASEARAAAPALPRGTASPGSQLAKAPKPKPESSAEQRLPLGNGIOT 183
Db 125 LPPRPPTKRKALEEPRAATPATLASKSVSSPSGSHLTKPRKPPDGNLESOHLPLGNGIOT 184
QY 184 MSASVORAVAMSSGDVPGARAGAVEGILIQOVFESGSKKCIQVGEFYTSPSKFED-SGSG 242
Db 185 MAASVORAVAVASGDVPGTGAVEGILIQOVFESGSKKCIQVGEFYTSPSKFEDPSGNL 244
QY 243 KKKARSSSGPKPLVRAKGAOGAARPGGEARLGQGSVPAPALPSPDQLHOKNDECAVC 302
Db 245 KKKARSSSLKLPVRAKGAOGR----DEQKVGQCCVPLPSLSEPOVNO-NEDECAVC 299
QY 303 RDGELICDGCPRAFHLACLSPPLREIPSGTWRCSCLQATVOEVPRAEPRPOEPPV 362
Db 300 HDGELICDGCPRAFHLACLSPPLREIPSGTWRCSCLQGRVQONLSQPEVSRPELPA 359
QY 363 ETPPLPGLRSAGEVRGPRGPELAGMDTTLVYKHLPAAPSAPLPGLDSSALHPLLCVGP 422
Db 360 ETPILVGLRSASEKTRGPRRELKASSDAAYTYVNLPAHPAAPL--LEPSALCPILLSAGN 417
QY 423 EGOONLAPGARCGVCGDGTDLVLRCTHCAAFHWRCHFPAGTSRPGTGRCSSGSDVTPA 482
Db 422 EGRPGAPARSVCYCGDGTDLVLRCAHCAAFHWRCHFPAGTSRPGTGRCSSGSDSTPT 477
QY 483 P-VEGVLP-SPARLAPGPAK--DDTASHEPALHRDLSLSEHTFEDGILQMAIOSMAR 538
Db 478 PGTBEAVPTSGPRAPAGLAKVGDSDASHDPVLRHDDLSLSEHTFEDGILQMAIOSMSR 537
QY 539 PAA---PPPS 545
Db 538 PLAEPTPPSS 547

```

RESULT 6
Q9JLM7 PRELIMINARY: PRT: 493 AA.
AC Q9JLM7:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE Autoimmune regulator.
GN AIRE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6.
RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
RT "Expression and alternative splicing of the mouse autoimmune regulator gene (Aire).";
DR EMBL: AF128117; AAF36462.1; -;
DR MGD: MGI:1338603; Aire.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000770; SAND domain.
DR InterPro: IPR004865; SP100.
DR InterPro: IPR001965; Znf_PHD.
DR InterPro: IPR001841; Znf_fing.
DR Pfam: PF00628; PHD; 2.

RESULT 8

09JLM6 PRELIMINARY: PRT: 492 AA.
 ID 09JLM6
 AC 09JLM6
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Autoimmune regulator.
 GN AIRE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6;
 RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
 RT "Expression and alternative splicing of the mouse autoimmune regulator
 gene (Aire).";
 RL Submitted (FEb-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF128119; AAF36464.1; -;
 DR MGD: MGI:1338803; Aire.
 DR InterPro: IPR000345; CytC_heme_bind.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000770; SAND_domain.
 DR InterPro: IPR004865; Sp100.
 DR InterPro: IPR001965; ZnF_PHD.
 DR Pfam: PF00628; PHD; 2.
 DR Pfam: PF03142; SAND; 1.
 DR Pfam: PF03172; Sp100; 1.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00184; RING; 2.
 DR SMART: SM00258; SAND; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 SQ SEQUENCE 492 AA; 52868 MW; 9BE6428E275E5781 CRC64;

Query Match

Best Local Similarity 63.3%; Score 1838; DB 11; Length 492;
 Matches 366; Conservative 31; Mismatches 83; Indels 70; Gaps 8;

QY 4 DAALRLRLRLHRTETIAVAANDSAFPLHLADHDVVPEDKFOETLHLKEGCPQAFHALL 63
 DB 5 DGMRLRLRLHRTETIAVAANDSAFPLHLADHDVVPEDKFOETLHLKEGCPQAFHALL 64
 QY 64 SWLLTODSTALIDFWRVLEFKDYNLERGRLOPLIDSPKDVDSLQPKRGRKPPAVKALY 123
 DB 65 SWLLTRDSGALIDFWRILFKDYNLERGRLOPLIDSPKDVDSLQPKRGRKPPAVKALY 124
 QY 124 PPPRLPTKRKASEAARAAPALTPRGTSAPSGOLAKPKPKKPESSAEQORLPLGNGIOT 183
 DB 125 LPPRPPTKRKALEEPRAATPATLASKSVSSPSGSHLTKTPPKPDGNGLESQHLPLGNGIOT 184
 QY 184 MSASVQRAVAMSSGDVPGARGAVEGILIQVFESGSKKCIQVGGEEYTPSKFED-SCSG 242
 DB 185 MAASVQRAVAVASGDVPGTRGAVEGILIQVFESGSKKCIQVGGEEYTPSKFEDPSGNL 244
 QY 243 KKKARSSSGPKPLVRAKGAOGAAPGGEGEARLGOQGSVPAPLAPSPDLOHKNEDCAV 302
 DB 245 KKKARSSSGSLKPYVRAKGAQAVITIPGRDEQKVGQCGVPLPLSPSEPOVNG-NEDECAV 303
 QY 303 RDGGLLICCDGCPRAFHALLCLSPRLREISGTWRSCSCLOATVOEVOAPRAEPPRPOEPV 362
 DB 304 HDGGLLICCDGCPRAFHALLCLSPRLREISGTWRSCSCLOATVOEVOAPRAEPPRPOEPV 363
 QY 363 ETPLEPGLRSAGEEVRGPGEPLAGMDTTLVKKHLPAFPAAPLPLGLDSSALHPLLCVGP 422
 DB 364 ETP-----GP 368
 QY 423 EGGONLAPGARCGCGDGLVLRCTHCAAFHMRCHFPAGTSRPGTGLRCRSCSGDVTPTA 482

DB 369 -----APSARCSYCGDTEVLRCAHCAAFHMRCHFPPTAARPGTNLRKSCSADSTPT 422
 QY 483 P-VEGVLP--SPARLAPGPAK--DDTASHEPALHRDDSELSSETPFDGILOMATQSMAR 538
 DB 423 PGTPGEAVPTSGPRPAGLAKVGDSDASHDVLHRDDLESLLNHSFPGILOMATQSMAR 482
 QY 539 PAA---PPPS 545
 DB 483 PLAETPPFSS 492

RESULT 9

09JLM4 PRELIMINARY: PRT: 488 AA.

ID 09JLM4
 AC 09JLM4
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Autoimmune regulator.
 GN AIRE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6;
 RA Ruan Q.G., Wang C.Y., Shi J.D., She J.X.;
 RT "Expression and alternative splicing of the mouse autoimmune regulator
 gene (Aire).";
 RL Submitted (FEb-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF128121; AAF36466.1; -;
 DR MGD: MGI:1338803; Aire.
 DR InterPro: IPR000345; CytC_heme_bind.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000770; SAND_domain.
 DR InterPro: IPR004865; Sp100.
 DR InterPro: IPR001965; ZnF_PHD.
 DR Pfam: PF00628; PHD; 2.
 DR Pfam: PF03172; Sp100; 1.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00249; PHD; 2.
 DR SMART: SM00184; RING; 2.
 DR SMART: SM00258; SAND; 1.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 SQ SEQUENCE 488 AA; 52457 MW; 628EDC8A863C7355 CRC64;

Query Match

Best Local Similarity 62.9%; Score 1824; DB 11; Length 488;
 Matches 365; Conservative 31; Mismatches 80; Indels 74; Gaps 9;

QY 4 DAALRLRLRLHRTETIAVAANDSAFPLHLADHDVVPEDKFOETLHLKEGCPQAFHALL 63
 DB 5 DGMRLRLRLHRTETIAVAANDSAFPLHLADHDVVPEDKFOETLHLKEGCPQAFHALL 64
 QY 64 SWLLTODSTALIDFWRVLEFKDYNLERGRLOPLIDSPKDVDSLQPKRGRKPPAVKALY 123
 DB 65 SWLLTRDSGALIDFWRILFKDYNLERGRLOPLIDSPKDVDSLQPKRGRKPPAVKALY 124
 QY 124 PPPRLPTKRKASEAARAAPALTPRGTSAPSGOLAKPKPKKPESSAEQORLPLGNGIOT 183
 DB 125 LPPRPPTKRKALEEPRAATPATLASKSVSSPSGSHLTKTPPKPDGNGLESQHLPLGNGIOT 184
 QY 184 MSASVQRAVAMSSGDVPGARGAVEGILIQVFESGSKKCIQVGGEEYTPSKFED-SCSG 242
 DB 185 MAASVQRAVAVASGDVPGTRGAVEGILIQVFESGSKKCIQVGGEEYTPSKFEDPSGNL 244
 QY 243 KKKARSSSGPKPLVRAKGAOGAAPGGEGEARLGOQGSVPAPLAPSPDLOHKNEDCAV 302
 DB 245 KKKARSSSGSLKPYVRAKGAQAVITIPGRDEQKVGQCGVPLPLSPSEPOVNG-NEDECAV 303

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:45:22 ; Search time 38 Seconds

(without alignments)
1911.094 Million cell updates/sec

Title: US-09-509-595B-2

Perfect score: 2902

Sequence: 1 MARDALRLRLRLHRTETAV.....DGILOAIQSNAPAPPPS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDSL/gcgdata/geneseq/geneeqp-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT:*
10: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT:*
11: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT:*
12: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT:*
13: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT:*
14: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT:*
15: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT:*
16: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT:*
17: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT:*
18: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT:*
19: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT:*
20: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT:*
21: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT:*
22: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT:*
23: /SIDSL/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2902	100.0	545	20	AAV06073 Human autoimmune p
2	2902	100.0	545	20	AAV01712 A human autoimmune
3	2005	69.1	552	20	AAV06074 Mouse AIRP protein
4	1423.5	49.1	348	20	AAV01713 A human autoimmune
5	503.5	17.4	254	20	AAV01714 A human autoimmune
6	241	8.3	1911	17	AAAG9534 Dermatomyositis sp
7	241	8.3	1912	22	AAAG39288 Human polypeptide
8	238	8.2	1936	22	AAAM1074 Human polypeptide
9	236.5	8.1	689	23	ABR07636 Human speckled 110
10			351	13	AAK22280 M1-2 antigen clone

11	227	7.8	884	22	ABR71641 Drosophila melanog
12	227	7.6	1050	22	AAAG3902 Human polypeptide,
13	220.5	7.6	753	23	ABR07637 Human speckled 140
14	213.5	7.4	753	19	AAAG5747 Leukocyte specific
15	213.5	7.4	771	22	ABG11257 Novel human diagno
16	212	7.3	663	21	AAV71210 Human irritable bo
17	210.5	7.3	245	22	AAAG4440 Human protein sequ
18	204.5	7.0	885	22	ABG11256 Novel human diagno
19	203	7.0	759	22	AAAG4613 Human protein sequ
20	192	6.6	582	22	AAAG1852 Human polypeptide
21	192	6.6	582	22	AAAG1853 Human polypeptide
22	191	6.6	628	22	AAAG0067 Human polypeptide
23	187.5	6.5	898	22	ABR60194 Drosophila melanog
24	186.5	6.4	1139	22	ABR61347 Drosophila melanog
25	186	6.4	504	21	AAAG2081 Human ORFX ORF1845
26	184	6.3	335	21	AAAG2612 Human ORFX ORF2376
27	181	6.2	883	22	ABR17771 Drosophila melanog
28	179.5	6.2	319	22	AAAG74399 Human colon cancer
29	179	6.2	605	22	ABG18143 Novel human diagno
30	177	6.1	1413	23	AAE21729 Human PKIN-24 prot
31	175.5	6.0	581	22	AAAG4066 Human polypeptide
32	175.5	6.0	634	22	AAAG4512 Human protein sequ
33	169	5.8	334	23	ABR89370 Human polypeptide
34	168	5.8	1134	22	AAAG0064 Human protein tyro
35	168	5.8	1570	23	ABR7344 Novel human protei
36	167.5	5.8	728	21	AAAG2035 Arabidopsis thalia
37	167.5	5.8	777	21	AAAG2036 Arabidopsis thalia
38	167.5	5.8	1080	21	AAAG31542 Arabidopsis thalia
39	167.5	5.8	1092	21	AAAG31541 Arabidopsis thalia
40	167.5	5.8	1138	21	AAAG31540 Arabidopsis thalia
41	166.5	5.7	594	21	AAAG2037 Arabidopsis thalia
42	166	5.7	549	22	AAU16301 Human novel secret
43	165	5.7	489	22	ABG14285 Novel human diagno
44	164.5	5.7	393	21	AAAG28067 Human secreted pro
45	164.5	5.7	393	21	AAAG28068 Human secreted pro

ALIGNMENTS

RESULT 1	AAV06073	standard; Protein: 545 AA.
ID	AAV06073	
XX	AAV06073	
AC	AAV06073	
XX	16-AUG-1999	(first entry)
XX		
DE	Human autoimmune polyglandular disease type 1 (APGDI) protein.	
XX		
XX	Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy:	
KW	ABPCECD: autoimmune polyglandular disease type 1; APGDI; AIRP;	
KW	transcription factor; autoimmune disease; diagnosis; gene therapy;	
KW	human.	
XX		
XX	Homo sapiens.	
OS		
FT	Key	Location/Qualifiers
FT	Region	7..11
FT		/note- "LXXLL motif located in putative helical region, signature for nuclear receptor binding"
FT	Region	113..114
FT		/note- "putative bi-partite nuclear localisation"
FT	Region	131..133
FT		/note- "putative bi-partite nuclear localisation"
FT	Domain	299..340
FT		/note- "PHD zinc finger domain"
FT	Domain	434..475
FT		/note- "PHD zinc finger domain"
XX		
PN	W09918197-A2.	

PD 15-APR-1999.
 XX
 PF 02-OCT-1998; 98WO-EP06294.
 XX
 PR 12-NOV-1997; 97EP-0119810.
 PR 02-OCT-1997; 97EP-0117154.
 PR 08-OCT-1997; 97EP-0117398.
 XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (NAPU-) NAT PUBLIC HEALTH INST.
 XX
 PI Aaltonen J, Björnsen P, Horelli-Kultunen N, Lehrach H;
 PI Palotie A, Peltonen L, Perheentupa J, Yaspo M;
 XX
 DR WPI: 1999-287735/24.
 DR N-PSDB; AAX58605.
 XX
 PT New polypeptide which co-segregates in mutated form - with
 PT autoimmune polyendocrinopathy candidiasis ectodermal dystrophy
 XX
 PS Claim 1; Fig 2A; 77pp; English.
 XX
 CC This is the amino acid sequence of human autoimmune polyglandular
 CC disease type 1 (APGD1 or AIRE), as predicted from a composite
 CC sequence (see AAX58605) of isolated cDNA clones and a PCR extension
 CC product. APGD1 protein is a transcription factor or transcription-
 CC associated factor that may associate with vimentin fibres, perhaps
 CC as part of a docking mechanism regulating nuclear translocation.
 CC Aggregates of the mutated protein may prevent formation of vimentin
 CC intermediate filaments. Mutated APGD1 polypeptides co-segregate
 CC with autoimmune polyendocrinopathy candidiasis ectodermal dystrophy
 CC (APECED). The invention provides vectors and host cells for
 CC preparation of APGD1 polypeptides. A claimed method for testing
 CC for carrier status of APECED, or for the disease state, involves
 CC testing for a mutation in the APGD1 gene or for a mutated form of
 CC the APGD1 polypeptide.
 XX
 SQ Sequence 545 AA;
 Query Match 100.0%; Score 2902; DB 20; Length 545;
 Best Local Similarity 100.0%; Pred. No. 1.5e-211;
 Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MATDALRLRLRLHRTETAVAVDSAPFLHALLADHDVPEDEKFOETLHKEEGCQAFH 60
 DB 1 MATDALRLRLRLHRTETAVAVDSAPFLHALLADHDVPEDEKFOETLHKEEGCQAFH 60
 QY 61 ALLSWLLTQDSTALIDFWRVLFKDYNLERYGRLQPIIDSFPKDVLDSQPRKGRPPAVPK 120
 DB 61 ALLSWLLTQDSTALIDFWRVLFKDYNLERYGRLQPIIDSFPKDVLDSQPRKGRPPAVPK 120
 QY 121 ALVPPRLPTPKRKAASEEAAAPALTPRGTAASPSQKAKPKPKPESAEQORLPLNG 180
 DB 121 ALVPPRLPTPKRKAASEEAAAPALTPRGTAASPSQKAKPKPKPESAEQORLPLNG 180
 QY 181 IQTMSAVORAVAMSSGDVVGARGAVGILIQVFESGSKCQVGGSEFTPSKFEDESG 240
 DB 181 IQTMSAVORAVAMSSGDVVGARGAVGILIQVFESGSKCQVGGSEFTPSKFEDESG 240
 QY 241 SGKTKRSSSGRPPLVRKAGAGAPGEGARLQGQGSVPAPLALPSDPOLHQNEDCA 300
 DB 241 SGKTKRSSSGRPPLVRKAGAGAPGEGARLQGQGSVPAPLALPSDPOLHQNEDCA 300
 QY 301 VCRDGGELICCGCCPRAFLACISPLRLRETSGTWRCSSCLQATVQEVQRAEPPRQEP 360
 DB 301 VCRDGGELICCGCCPRAFLACISPLRLRETSGTWRCSSCLQATVQEVQRAEPPRQEP 360
 QY 361 FVEYPLPGLRSAGEVRPPEGLAGMDTLLVYKHLPAAPSAAPLPGDSSALHPLCY 420
 DB 361 FVEYPLPGLRSAGEVRPPEGLAGMDTLLVYKHLPAAPSAAPLPGDSSALHPLCY 420
 QY 421 GPEGQONLAPGARGGVCGDGTDLRCTHCAAFHWRCHPEPAGTSRPGTGLRSCSGDYT 480
 DB 421 GPEGQONLAPGARGGVCGDGTDLRCTHCAAFHWRCHPEPAGTSRPGTGLRSCSGDYT 480

DB 421 GPEGQONLAPGARGGVCGDGTDLRCTHCAAFHWRCHPEPAGTSRPGTGLRSCSGDYT 480
 QY 481 PAPVEGLAPSPARLAPGPAKDOTASHEPALRDDELSLEHTFPGCIIQMAIOSMARPA 540
 DB 481 PAPVEGLAPSPARLAPGPAKDOTASHEPALRDDELSLEHTFPGCIIQMAIOSMARPA 540
 QY 541 APFES 545
 DB 541 APFES 545
 RESULT 2
 ID AAY01712 standard; Protein; 545 AA.
 XX
 AC AAY01712;
 XX
 DT 24-JUN-1999 (first entry)
 XX
 DE A human autoimmune regulator-1 (AIRE-1) protein.
 XX
 KW Autoimmune regulator-1; AIRE-1; immune maturation; immune response;
 KW disease; autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;
 KW APECED; autoimmune polyglandular syndrome type 1; APS I.
 OS Homo sapiens.
 XX
 PN WO915559-A1.
 XX
 PD 01-APR-1999.
 XX
 PF 23-SEP-1998; 98WO-FI00749.
 XX
 PR 23-SEP-1997; 97FI-0003762.
 XX
 PA (FIIM-) FINNISH IMMUNOTECHNOLOGY LTD.
 XX
 PI Antonarakis S, Helio M, Krohn K, Kudoh J, Laitola M;
 PI Peterson P, Scott H, Shimizu N;
 XX
 DR WPI: 1999-244390/20.
 DR N-PSDB; AAX26936.
 XX
 PT Autoimmune regulator 1 (AIRE) DNA sequence
 XX
 PS Claim 5; Page 24-26; 59pp; English.
 XX
 CC The present sequence represents an autoimmune regulator-1 (AIRE-1)
 CC protein. The AIRE polypeptides and polynucleotides can be used
 CC in methods for the diagnosis and treatment of diseases related
 CC to immune maturation and regulation of immune response towards
 CC self and nonself. They can be used particularly in the diagnosis
 CC and treatment of autoimmune polyendocrinopathy candidiasis
 CC ectodermal dystrophy (APECED) (also known as autoimmune
 CC polyglandular syndrome type 1 (APS I)).
 XX
 SQ Sequence 545 AA;
 Query Match 100.0%; Score 2902; DB 20; Length 545;
 Best Local Similarity 100.0%; Pred. No. 1.5e-211;
 Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MATDALRLRLRLHRTETAVAVDSAPFLHALLADHDVPEDEKFOETLHKEEGCQAFH 60
 DB 1 MATDALRLRLRLHRTETAVAVDSAPFLHALLADHDVPEDEKFOETLHKEEGCQAFH 60
 QY 61 ALLSWLLTQDSTALIDFWRVLFKDYNLERYGRLQPIIDSFPKDVLDSQPRKGRPPAVPK 120
 DB 61 ALLSWLLTQDSTALIDFWRVLFKDYNLERYGRLQPIIDSFPKDVLDSQPRKGRPPAVPK 120
 QY 121 ALVPPRLPTPKRKAASEEAAAPALTPRGTAASPSQKAKPKPKPESAEQORLPLNG 180
 DB 121 ALVPPRLPTPKRKAASEEAAAPALTPRGTAASPSQKAKPKPKPESAEQORLPLNG 180

```

OY 181 IOTMSAVORAVAMSSGDVPGARGAVEGILIOVFESGSKCICQVGGERTYPSKFEFDSG 240
DB 181 IOTMSAVORAVAMSSGDVPGARGAVEGILIOVFESGSKCICQVGGERTYPSKFEFDSG 240
OY 241 SGKNRARSSSGPKPLVRAKGAOGAARPGGEGARLGQGSVPAPLALSDPOLHOKNEDECA 300
DB 241 SGKNRARSSSGPKPLVRAKGAOGAARPGGEGARLGQGSVPAPLALSDPOLHOKNEDECA 300
OY 301 VCRDGGELICDGCPRAFHACLSPLEIREIPSGTWRCSSCLQATVQEVOPRAEPRPOEP 360
DB 301 VCRDGGELICDGCPRAFHACLSPLEIREIPSGTWRCSSCLQATVQEVOPRAEPRPOEP 360
OY 361 PVEPPLPGLRSAGEVRNRPPEPEPLAGMDTTLVYKHLPAAPSAAPLPGLDSSALHPLCY 420
DB 361 PVEPPLPGLRSAGEVRNRPPEPEPLAGMDTTLVYKHLPAAPSAAPLPGLDSSALHPLCY 420
OY 421 GPEGQONLAPGARGCGVCGDGTDLRCTHCAAFHMRCHFPAGTSRPGTGLRCRSGDVT 480
DB 421 GPEGQONLAPGARGCGVCGDGTDLRCTHCAAFHMRCHFPAGTSRPGTGLRCRSGDVT 480
OY 481 PAPVEGVLAAPSPARLAPGPAKDDTASHEPALHRDLESLSLSEHTFDGILQMATQSMARPA 540
DB 481 PAPVEGVLAAPSPARLAPGPAKDDTASHEPALHRDLESLSLSEHTFDGILQMATQSMARPA 540
OY 541 APFPS 545
DB 541 APFPS 545

RESULT 3
AA06074
ID AA06074 standard; Protein: 552 AA.
AC AA06074;
XX
XX 16-AUG-1999 (first entry)
DE Mouse AIRE protein.
XX
XX Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;
KW APECED; autoimmune polyglandular disease type 1; APGD; AIRE;
KM transcription factor; autoimmune disease; diagnosis; gene therapy;
KM mouse; animal model.
XX
OS Mus sp.
XX
XX MO9918197-A2.
XX
XX 15-APR-1999.
XX
XX 02-OCT-1998; 98WO-EP06294.
XX
XX 12-NOV-1997; 97EP-0119810.
XX 02-OCT-1997; 97EP-0117154.
XX 08-OCT-1997; 97EP-0117398.
XX
XX (PLAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.
XX (NAPU-) NAT PUBLIC HEALTH INST.
XX
XX Aallonen J, Bjorres P, Horelli-Kultunen N, Lehnach H;
XX PI Palotie A, Peltonen L, Perheentupa J, Yaspo M;
XX DR N-PSDB; AAX58606.
XX
XX WPI; 1999-287735/24.
XX
XX New polypeptide which co-segregates in mutated form - with
XX PT autoimmune polyendocrinopathy candidiasis ectodermal dystrophy
XX PS Example 20; Fig 14A-C; 77pp; English.
XX
XX This is the amino acid sequence of murine AIRE, as deduced from
XX CC an isolated cDNA clone (see AAX58606). Murine AIRE is the homologue

```

```

CC of the human AIRE (or autoimmune polyglandular disease type 1,
CC APECED) polypeptide (see AA06073). The overall identity between the
CC mouse and human AIRE proteins is 72.37%. Human mutated APGD1
CC co-segregates with autoimmune polyendocrinopathy candidiasis
CC ectodermal dystrophy (APECED). The murine homologue may be used to
CC develop an animal model for APECED, to examine the events that lead
CC to the development of APECED and possibly to develop agents for
CC preventing and/or treating this autoimmune disease.
XX
XX Sequence 552 AA:
SQ
Query Match 69.1%; Score 2005; DB 20; Length 552;
Best Local Similarity 71.8%; Pred. No. 1.5e-143;
Matches 395; Conservative 37; Mismatches 108; Indels 10; Gaps 6;
OY 4 DAALRLRLRLRTEAVAVDSAFPLHLADHDVPEKFEQETLHKKEGCPQAFHALL 63
DB 5 DGMRLRLRLRHTETAVVADSAFPLHLADHDVPEKFEQETLHKKEGCPQAFHALL 64
OY 64 SWLLTODSTATILDPMRVLFKDYNLERYGRLOPILDSFPKDVLSOPRKGRKPPAVKALV 123
DB 65 SWLLTRDGAITLDFWRLIFKDYNLERYGRLOPILDSFPKDVLSOPRKGRKPPAVKALV 124
OY 124 PPRPLPTKRKASEARAAAPALPRGTASPGSOLKAPPKKPESSAEQRLPLNGIQT 183
DB 125 LPPRPPTKRKALEBRATPPATLASKSVSPSHLTKPKPKKPDGNLESQHLPLNGIQT 184
OY 184 MSASVORAVAMSSGDVPGARGAVEGILIOVFESGSKCICQVGGERTYPSKFEFDSG 242
DB 185 MAASVORAVAVASGDVPGARGAVEGILIOVFESGSKCICQVGGERTYPSKFEFDSG 244
OY 243 KNRARSSSGPKPLVRAKGAOGAARPGGEGARLGQGSVPAPLALSDPOLHOKNEDECAVC 302
DB 245 KNRARSSSGPKPLVRAKGAOGAARPGGEGARLGQGSVPAPLALSDPOLHOKNEDECAVC 304
OY 303 RDGELICDGCPRAFHACLSPLEIREIPSGTWRCSSCLQATVQEVOPRAEPRPOEPV 362
DB 305 HDGELICDGCPRAFHACLSPLEIREIPSGTWRCSSCLQATVQEVOPRAEPRPOEPV 364
OY 363 ETPPLPGLRSAGEVRNRPPEPEPLAGMDTTLVYKHLPAAPSAAPLPGLDSSALHPLCY 422
DB 365 ETPPLPGLRSAGEVRNRPPEPEPLAGMDTTLVYKHLPAAPSAAPLPGLDSSALHPLCY 422
OY 423 EGOONLAPGARGCGVCGDGTDLRCTHCAAFHMRCHFPAGTSRPGTGLRCRSGDVT 482
DB 423 EGRGPAAPSAARCSVCGDGTDLRCTHCAAFHMRCHFPAGTSRPGTGLRCRSGDVT 482
OY 483 P-VEGVLAAP-SPARLAPGPAK-DDTASHEPALHRDLESLSLSEHTFDGILQMATQSMAR 538
DB 483 PGTGGEAVPTSGPRAPGLAKVGDSDASHDPVLAHRDLESLSLSEHTFDGILQMATQSMAR 542
OY 539 PAA---PFPS 545
DB 543 PLAEPTPFSS 552

RESULT 4
AA01713
ID AA01713 standard; Protein: 348 AA.
XX
XX AA01713;
XX
XX 24-JUN-1999 (first entry)
XX
XX A human autoimmune regulator-2 (AIR-2) protein.
XX
XX Autoimmune regulator-2; AIR-2; Immune maturation; Immune response;
XX KW disease; autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;
XX KW APECED; autoimmune polyglandular syndrome type I; APS I.
XX
XX Homo sapiens.
XX OS
XX PN WO9915559-A1.

```

XX 01-APR-1999.
 PD 23-SEP-1998; 98WO-F100749.
 XX 23-SEP-1997; 97FI-0003762.
 PR (FIIM-) FINNISH IMMUNOTECHNOLOGY LTD.
 PA Antonarakis S, Helmo M, Krohn K, Kudoh J, Lalioti M;
 PI Peterson P, Scott H, Shimizu N;
 XX WPI: 1999-244390/20.
 DR N-PSDB; AAX26937.
 XX
 PT Autoimmune regulator 1 (AIR1) DNA sequence
 PS Claim 7; Page 29-30; 59pp; English.
 XX
 CC The present sequence represents an autoimmune regulator-2 (AIR-2)
 CC protein. The AIR polypeptides and polynucleotides can be used
 CC in methods for the diagnosis and treatment of diseases related
 CC to immune maturation and regulation of immune response towards
 CC self and nonself. They can be used particularly in the diagnosis
 CC and treatment of autoimmune polyendocrinopathy candidiasis
 CC ectodermal dystrophy (APECED) (also known as autoimmune
 CC polyglandular syndrome type I (APS I)).
 CC
 SO Sequence 348 AA;

Query Match 49.1%; Score 1423.5; DB 20; Length 348;
 Best Local Similarity 83.5%; Pred. No. 9.7e-100;
 Matches 269; Conservative 10; Mismatches 38; Indels 5; Gaps 3;

QY 227 GGEYTPSKFEDSGSGKNKRRSSGPKPLVRAKAGAGAPG---GGEARLGGQGSVPAPL 283
 DB 29 GGCWMPDWMGTGGGKISGPGSMAGQRLGSSGTQRCWCSGCFKEVAL--RRVLHPSFV 87
 QY 284 ALPSPDLHOKNDECAVCRDGGELICDCCPRAFHLACSLPPLREIPSGTWRCSSCLQA 343
 DB 88 CM-GVSCLOKNEDECAVCRDGGELICDCCPRAFHLACSLPPLREIPSGTWRCSSCLQA 146
 QY 344 TVQEVQRAEPPRPQEPVETPLPPLRLSAGEEVRGPPGEPPLAGMDTTLVYKHLPPAPSA 403
 DB 147 TVQEVQRAEPPRPQEPVETPLPPLRLSAGEEVRGPPGEPPLAGMDTTLVYKHLPPAPSA 206
 QY 404 ALPGLDSSALHPLLCVGPBGQONLAPGARGVCGDGTDLVLRCTHCAAFHMRCHFPAGT 463
 DB 207 ALPGLDSSALHPLLCVGPBGQONLAPGARGVCGDGTDLVLRCTHCAAFHMRCHFPAGT 266
 QY 464 SRPGLRCRSCSGDYTPAPEGVLAAPSPARLAPGPAKDDTASHHPALHRDLSLSEH 523
 DB 267 SRPGLRCRSCSGDYTPAPEGVLAAPSPARLAPGPAKDDTASHHPALHRDLSLSEH 326
 QY 524 TPDGIIQMAIQSMARPAAPPS 545
 DB 327 TEDGIIQMAIQSMARPAAPPS 348

RESULT 5
 AAY01714
 ID AAY01714 standard; Protein; 254 AA.
 XX
 AC AAY01714;
 XX
 DT 24-JUN-1999 (first entry)
 XX
 DE A human autoimmune regulator-3 (AIR-3) protein.
 XX
 KW Autoimmune regulator-3; AIR-3; immune maturation; immune response;
 KW disease; autoimmune polyendocrinopathy candidiasis ectodermal dystrophy;
 APCECD; autoimmune polyglandular syndrome type I; APS I.
 XX

OS Homo sapiens.
 XX
 PN W09915559-A1.
 XX
 PD 01-APR-1999.
 XX
 PE 23-SEP-1998; 98WO-F100749.
 PR 23-SEP-1997; 97FI-0003762.
 XX
 PA (FIIM-) FINNISH IMMUNOTECHNOLOGY LTD.
 PI Antonarakis S, Helmo M, Krohn K, Kudoh J, Lalioti M;
 PI Peterson P, Scott H, Shimizu N;
 XX WPI: 1999-244390/20.
 DR N-PSDB; AAX26938.
 XX
 PT Autoimmune regulator 1 (AIR1) DNA sequence
 PS Claim 7; Page 32-33; 59pp; English.
 XX
 CC The present sequence represents an autoimmune regulator-3 (AIR-3)
 CC protein. The AIR polypeptides and polynucleotides can be used
 CC in methods for the diagnosis and treatment of diseases related
 CC to immune maturation and regulation of immune response towards
 CC self and nonself. They can be used particularly in the diagnosis
 CC and treatment of autoimmune polyendocrinopathy candidiasis
 CC ectodermal dystrophy (APECED) (also known as autoimmune
 CC polyglandular syndrome type I (APS I)).
 CC
 SO Sequence 254 AA;

Query Match 17.4%; Score 503.5; DB 20; Length 254;
 Best Local Similarity 43.8%; Pred. No. 3.5e-30;
 Matches 126; Conservative 17; Mismatches 76; Indels 69; Gaps 11;

QY 227 GGEYTPSKFEDSGSGKNKRRSSGPKPLVRAKAGAGAPG---GGEARLGGQGSVPAPL 283
 DB 29 GGCWMPDWMGTGGGKISGPGSMAGQRLGSSGTQRCWCSGCFKEVAL--RRVLHPSFV 87
 QY 284 ALPSPDLHOKNDECAVCRDGGELICDCCPRAFHLACSLPPLREIPSGTWRCSSCLQA 343
 DB 88 CM-GVSCLOKNEDECAVCRDGGELICDCCPRAFHLACSLPPLREIPSGTWRCSSCLQA 146
 QY 344 TVQEVQRAEPPRPQEPVETPLPPLRLSAGEEVRGPPGEPPLAGMDTTLVYKHLPPAPSA 403
 DB 147 TVQEVQRAEPPRPQEPVETPLPPLRLSAGEEVRGPPGEPPLAGMDTTLVYKHLPPAPSA 192
 QY 404 ALPGLDSSALHPLLCVGPBGQONLAPGARGVCGDGTDLVLRCTHCAAFHMRCHFPAGT 463
 DB 193 -----KCVW-----VLRVSR-----MLLVAVAGC 212
 QY 464 SRPGLRCRSCSGDYTPAPEGVLAAPSPARLAPGPAKDDTASHHPA 510
 DB 213 AE-----MVRTCCGVTLAPLPSTGATSQPA--PPGPR--ACAADPA 251

RESULT 6
 AAR99534
 ID AAR99534 standard; Protein; 1911 AA.
 XX
 AC AAR99534;
 XX
 DT 30-OCT-1996 (first entry)
 XX
 DE Dermatomyositis specific autoantigen, Mi-2.
 XX
 KW Mi-2; autoantigen; collagen disease; chromosome 12; 12p13;
 KW helicase; dermatomyositis; diagnosis.
 XX
 OS Homo sapiens.
 XX

Key	Location/Qualifiers
Region	53..73
Region	/label= "Region_a"
Region	/note= "contains 3 potential core target motifs"
Region	113..133
Region	/label= "Region_b"
Region	/note= "contains 4 potential core target motifs"
Region	133..143
Region	/note= "possible electrostatic interaction with chromatin of histones"
Region	257..287
Region	/label= "Region_c"
Region	/note= "contains 3 potential core target motifs"
Region	747..758
Region	/note= "Helicase-specific motif I"
Region	782..793
Region	/note= "Helicase-specific motif IA"
Region	869..877
Region	/note= "Helicase-specific motif II"
Region	897..911
Region	/note= "Helicase-specific motif IIR"
Region	943..959
Region	/label= "Region_d"
Region	/note= "contains 1 potential core target motif"
Region	949..960
Region	/note= "Helicase-specific motif IV"
Region	1121..1144
Region	/note= "Helicase-specific motif V"
Region	1149..1183
Region	/note= "Helicase-specific motif VI"
Region	DE19509279-C1.
Region	15-MAY-1996.
Region	15-MAR-1995; 95DE-1009279.
Region	15-MAR-1995; 95DE-1009279.
Region	(PRIV-) PRIVATE INST IMMUNOLOGIE & MOLEKULARGEN.
Region	Renz M, Seelig HP;
Region	WPI. 1996-240280/25.
Region	N-PSDB: AAT32301.
Region	DNA encoding dermatomyositis specific autoantigen - useful for differential diagnosis and treatment of dermatomyositis
Region	Claim 1; Fig 2; 20pp; German.
Region	The present sequence is that of a 218 kD dermatomyositis specific auto-antigen, designated M1-2. The sequence numbering given in the specification starts at amino acid 2, i.e. the first Met residue is omitted. The protein is hydrophilic, acidic and protruding regions of the protein are characteristic of helicases. The gene corresponding to the cDNA (AAT32301) encoding M1-2, was localised to chromosome 12 (12p13). The DNA can be used for the recombinant production of M1-2 which is used for, e.g. the differential diagnosis of collagen diseases, esp. dermatomyositis, e.g. by immunoassay or Western blotting.
Region	Sequence 1911 AA:
Region	Query Match 8.3%; Score 241; DB 17; Length 1911;
Region	Best Local Similarity 23.1%; Pred. No. 3.3e-09;
Region	Matches 97; Conservative 51; Mismatches 172; Indels 100; Gaps 14
Region	129 PPKRAASEARAAAPALTPRGTAAGSGOLAKAPKPKPESSAEQRLPLGNGIQTMSASY 188
Region	216 PFGSSGASVAANAANAANVAVESMTVATEV-APPPEVPEPIKAKTKEGKGPNA----- 269
Region	189 QAAVAMSSDVGANGAVGICILIQVFESG-----SKKCIQVNGEYTTPEKFPD----- 238

Db	270	-RRRKGSRRVPDAKKPKRKVKVPLKIKLGGFGSKRRRSSSEDDDLVESDFDASINSY	328
Qy	229	SGSGGNKARSSSGPKPLVRAKGAQAAPGGEGEARLGGQGSVPAPLALPSPQLHOKNE	298
Db	329	SVSDGSTSRSSSRKKLRTTK-----KRRKGE-EVTAVDGETD-----HQDY	371
Qy	239	CAVCNDGGELICDDGPAPFAHLCLSPLEIRISGTRCSSCLQATQVEQVPAEPRQ	358
Db	372	CEVCGQGGGELLICDTCPRAYHMYCLDPDMKAEGRKSCPCKE---EGIQWEAKED---	425
Qy	359	EPFVETPLPPLGRASAEVRGPGEPGLAGDITLVYKHLPAAPSAAPLPGLDSSALHPL	418
Db	426	-----NSEEELIEEYGGDLEEDD-----HHMF-----	450
Qy	419	CVGPEGQQLAPGARCGVCGDGTDLVRLCTHCAAAFWHRCHEPACTSRPTGRLRCRSCGD	478
Db	451	-----CRCKDGGELLICDTCPPSSYHICLNPLPEIPINGEMLCPRTCP	495
Qy	419	VTPAVEGL-----APSPARL-AGCPAKDDTASHPALHRDLESLSHHTDGILO	531
Db	436	ALRGVKQKILIKWKGQPPSPTPVPRPDADPNTPSPKPLEGRERQPFVK---WQMSYX	552
RESULT 7			
ID	AAM39288	standard; Protein; 1912 AA.	
AC	AAM39288;		
XX	22-OCT-2001	(first entry)	
DE	Human polypeptide SEQ ID NO 2433.		
XX	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropsychic; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;		
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia.		
XX			
OS	Homo sapiens.		
XX	WO200153312-A1.		
PN	26-JUL-2001.		
PD			
XX	26-DEC-2000; 2000WO-US34263.		
PF			
XX	21-JAN-2000; 2000US-0488725.		
XX	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Dymnac RT;		
XX			
DR	WPI: 2001-442253/47.		
DR	N-PSDB; AAI58444.		
XX			
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
XX			
PS	Example 4; SEQ ID NO 2433; 10078bp; English.		
XX			
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and		
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,		

Db 518 ALKGVQKILMKMGQSPSPVPRPPADPTPSPKPLEGRERQFVX---WQMSYV 574

RESULT 9

ABB07636 ID ABB07636 standard; Protein: 689 AA.

AC ABB07636;

DT 20-MAY-2002 (first entry)

XX Human speckled 110 (Sp110) polypeptide.

XX Sp110, speckled 110, nuclear body; gene transcription; dimerisation;

KW nuclear hormone receptor; primary biliary cirrhosis; PBC; human;

KM cell differentiation; cellular defence; cytosolic; anorectic;

XX antiinflammatory.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 6..109 /note= "Sp100-like domain"

FT Domain 454..532 /note= "SAND domain"

FT Domain 537..577 /note= "plant homeobox domain"

FT Domain 606..674 /note= "bromodomain"

XX WO200208383-A2.

XX 31-JAN-2002.

XX 24-JUL-2001: 2001WO-US23248.

XX 24-JUL-2000: 2000US-220305P.

XX (GEHO) GEN HOSPITAL CORP.

XX Bloch DB, Bloch KD;

XX WPI: 2002-227042/28.

XX N-PSDB: ABA95062.

XX Novel substantially pure polypeptide component of the nuclear body,

XX speckled 110, useful in screening methods, in clinical diagnostic

XX methods, and for treating myeloid malignancies, inflammation and

XX obesity.

XX Claim 3; Fig 2; 55pp; English.

XX The invention relates to cloning and characterisation of a cDNA encoding

XX Sp110 (speckled 110), a novel 110 kDa polypeptide component of the

XX nuclear body. The Sp110 polypeptide comprises a Sp100-like domain, a

XX SAND domain, a plant homeobox domain, and a bromodomain. Sp110 functions

XX as an activator of gene transcription and serves as a nuclear hormone

XX receptor co-activator. Sp110 is useful in screening for identifying a

XX compound that modulates (i) Sp110 dimerisation; (ii) Sp110 binding to a

XX nuclear hormone receptor; (iii) binding of an Sp110 dimer to an Sp110-

XX binding nucleotide sequence and in a screening method for identifying a

XX polypeptide that dimerises with Sp110 to form a constitutively active or

XX hyperactive or inactive heterodimer. It is useful for diagnosing primary

XX biliary cirrhosis (PBC) and is also useful for producing Sp110-specific

XX antibodies, for inhibiting viral replication and facilitating

XX differentiation of cells, e.g. myeloid cells, and activation of cells

XX involved in host defence, to treat myeloid malignancies, to enhance

XX cellular defence mechanisms, to treat inflammation, to achieve

XX alteration in lipid profiles, to block estrogen receptors in treatment

XX of estrogen responsive tumours, and for treating obesity. The present

XX sequence represents the human Sp110 polypeptide.

SQ Sequence 689 AA:

Query Match 8.2%; Score 238; DB 23; Length 689;

Best Local Similarity 19.7%; Pred. No. 1,6e-09;

Matches 120; Conservative 44; Mismatches 137; Indels 308; Gaps 19;

QY 2 AFDALRLRLRLHRTETAVAVDSAPFLHLALADHDVPEKFOETLHKEKGC----- 55

DB 7 AMEEALFOHFMHQKLGIAVAIHKPFPEEGILDSITKRYMBSL-----EACNLLIPV 61

QY 56 PQAFHALLSWLLTQDSTAILDFWRLFKDYVLEKRGRLQPLIDSEPKVDLSQPRKGRK 115

DB 62 SRVYHNLITQL---ERTFNLSLVTFLSQVLRREYPLVTYIRSF-KRVGASIERQSRDT 117

QY 116 PAV-----PKALVPP----- 125

DB 118 PILLEAPYGLAEGSLHTPLALPPQPPQPCSPCAPVSEKGTSSQOODELSESPSS 177

QY 126 -PRLP-----TKRKASEARAAPAAALT----- 147

DB 178 DPVLPPLPALIQEGRSTSVTNDKLTSKMAEEDSEMPSLTSTVOVASDNLIPQIRKED 237

QY 148 ----- 147

DB 238 PQEMPHSPGSMPEIRONSPEPNDEPEQEVSTSPDKKKKKRCIMSTPKRHKKSL 297

QY 148 PRGTASP-GSOLAK-----PKKPSSA----- 170

DB 298 PRGTASSRHGIQKRLKRVDPQPKKDDSTCNSTVETRAQKARTECARKSRSEIIDTSE 357

QY 171 -----EGRRLPL-----GNGIO-----TMSASVORA-- 191

DB 358 MNEGKRSQKTPSTPRRTVQGAASPGHGIQELQVVDKYQKRDSDTWNSEVMNRKART 417

QY 192 -----VAMSSGDVPCARGAVEGILIQ 212

DB 418 KCAKRSKSEKKEKEDICSSSKRRFOKNIRGRKPSDVTDFCSKLPVTCGEAKGILYK 477

QY 213 QVESGSGSKKCIQ-VGGEFTYPSKFPEDSGGNKARSSGPRPLVRAKGAQGAAPGGEA 271

DB 478 KMKRHGSSVCKICRMDGTWLPNFEFEVEGKGRN---AKNWKNIKEG-----M 523

QY 272 RLGGGGSVPAPLALPSPDOLHOKNEDECAVCDGELICDDGCPRAFHILACTLSPPLREIP 331

DB 524 TLGE-----LKRKNSDECVCCGGGLCCGTCPRVFHECHLIPV-EAK 568

QY 332 SGTWRCSSC 340

DB 569 RMLWSCFRC 577

RESULT 10

AAAR22280 ID AAR22280 standard; Protein: 351 AA.

AC AAR22280;

DT 27-JUL-1992 (first entry)

XX M1-2 antigen clone L1 encoded protein.

XX Myositis-specific antigen; polymyositis scleroderma overlap;

XX syndrome; dermatomyositis; autoantibodies; immunodiagnostic assay.

XX Homo sapiens.

XX WO9204472-A.

XX 19-MAR-1992.

XX 05-SEP-1991: 91WO-US06418.

XX 07-SEP-1990: 90US-0579023.


```

QY 7 LRLRLRLHTETAVAVDSAFPLLHALADHDVPE---DKFOETLHLKEKGGCPQAFHALL 63
DB 42 IIRFFREKNEVEIASATIRPEFLMGLRDRSFISEQWEHQA--FRNLVPTVYVMCVL 99
QY 64 SMLLQDSTAILDFWVFLKRDYLERGRLQPLIDSF----- 100
DB 100 SEL---EKTFGMSHLEALSRINLMAYPDLEIYRSFQVNCYEHSPLOMNNVNDLEDRPR 156
QY 101 -----PKVDLSOPR-----GKRKPPAVPKA-----L 122
DB 157 LIPYKQENSNACHEMDITAVQEAALSSSPRCPESSSECEQALPKAGGDAEDAPSL 216
QY 123 VP-----PPRLP-----TKRKASE----- 136
DB 217 LIPVSKLAIQIDEGESEEMPKILPYDTEFTDLKTPQVYNEGPEKGLCLLPGEGERGSD 276
QY 137 -----EARAAPALTPRGTS-----PSQJKA-- 160
DB 277 DCEMCDGEERQEAASSLARSGSVSELENHPNNEGESEELASSLLYDNVPGAQSAVE 336
QY 161 -----KPPKPESSAE-----QORLPJGNG----- 180
DB 337 NEKSCVWCFSEEPVGSPEARTESDQACGTMDTVDIANNSTLGKPRKRRKRGHGMSSRM 396
QY 181 -IQTMASAVQRAVAMSSGDV-----PGAR----- 203
DB 397 RMRROKNSQONDNSKADQVVSSEKKANVNLKLSKIRGKRKGKPGTRFTQSDRAAQKV 456
QY 204 -----GAVGILLIQVFESGSKKCIQV--GGEFYTPSKFED 238
DB 457 RSRASRKHKDETVDKAPLLPVTCGVGKILHKKRLQOGLIVKCIOTEDGKWFTEPEFEI 516
QY 239 SGSGKNARSSSGPRPLVRAKGAOGAAGGGEAR--LGOOGSVAPPLAL----- 285
DB 517 KG---GHARSKNW--RLSVRC-----GGMPLRLMENGFLPDPPIRYRKRRKRLKSO 564
QY 286 ---PSDQLHOKNDECAVCRDGGELICDGCPRAFHILACLSPLREIPSGTWRCSSC-- 340
DB 565 NNSVDPCM--RNLDECEVCRDGELFCDCDSRVFHEHCHIPPV--EAERTPMNCIFCMA 621
QY 341 ----LQATVOEVQRAEPRPQE 359
DB 622 KESPGSQCCQESSEVLERQKCPPE 645

RESULT 14
AAW57747
ID AAW57747 standard; Protein; 753 AA.
AC AAW57747;
XX
DT 17-SEP-1998 (first entry)
XX
DE Leukocyte specific protein, Sp140.
XX
KW Sp140; leukocyte specific protein; gene transcription regulator; therapy;
  autoimmune disease; viral infection; cancer.
XX
OS Homo sapiens.
XX
PN W09814569-A1.
XX
PD 09-APR-1998.
XX
PE 02-OCT-1997; 97WO-US17715.
XX
PR 02-OCT-1996; 96US-0027347.
XX
PA (BLOC/) BLOC D B.
  (BLOC/) BLOC K D.
XX
PI Bloch DB, Bloch KD;
XX

```

```

DR WPI: 1998-286419/25.
DR N-PSDB: AAW24559.
XX
PT New isolated gene transcription regulator, Sp140 - used to develop
PT products for the diagnosis and treatment of auto-immune diseases,
PT viral infections or cancers
XX
PS Claim 11; Page 54-57; 81pp; English.
XX
CC This sequence represents the leukocyte specific protein, Sp140 of the
CC invention. The Sp140 polypeptides act as gene transcription regulators.
CC They can be used to develop products for use in the diagnosis and
CC treatment of autoimmune diseases such as primary biliary cirrhosis,
CC rheumatoid arthritis, systemic lupus erythematosus, Sjogren's syndrome,
CC scleroderma and multiple sclerosis, viral diseases including those caused
CC by herpes simplex virus, cytomegalovirus, HIV, hepatitis virus, human
CC T-cell leukaemia virus-1 (HTLV-1) and adenovirus, and cancers including
CC breast, ovary, prostate, bone, liver, pancreas or spleen, sarcomas and
CC melanomas.
XX
SQ Sequence 753 AA;

Query Match 7.4%; Score 213.5; DB 19; Length 753;
Best Local Similarity 20.7%; Pred. No. 1.3e-07;
Matches 129; Conservative 54; Mismatches 148; Indels 291; Gaps 26;

QY 9 LRLRLRLHTETAVAVDSAFPLLHALADHDVPE---DKFOETLHLKEKGGCPQAFHALLSW 65
DB 44 RFRREKNEVEIASATIRPEFLMGLRDRSFISEQWEHQA--FRNLVPTVYVMCVL 101
QY 66 LRLQDSTAILDFWVFLKRDYLERGRLQPLIDSF-----PKDV 104
DB 102 L---EKTFGMSHLEALSRINLMAYPDLEIYRSFQVNCYEHSPLOMNNVNDLEDRPRL 158
QY 105 -----DLSDPRK-----GKRPP-----ATPKA-----LVP 124
DB 159 PYKQENSNACHEMDITAVQEAALSSSPRCPESSSECEQALPKAGGDAEDAPSLP 218
QY 125 -----PPRLP-----TKRKASE----- 136
DB 219 VSKLAIQIDEGESEEMPKILPYDTEFTDLKTPQVYNEGPEKGLCLLPGEGERGSDC 278
QY 137 -----EARAAPALTPRGTS-----PSQJKA-- 160
DB 279 SEMCDGEERQEAASSLARSGSVSELENHPNNEGESEELASSLLYDNVPGAQSAVE 338
QY 161 -----KPPKPESSAE-----QORLPJGNG-----I 181
DB 339 KCSVCWCFSEEPVGSPEARTESDQACGTMDTVDIANNSTLGKPRKRRKRGHGMSSRM 398
QY 182 QTMASAVQRAVAMSSGDV-----PGAR----- 203
DB 399 RROKNSQONDNSKADQVVSSEKKANVNLKLSKIRGKRKGKPGTRFTQSDRAAQKVR 458
QY 204 -----GAVGILLIQVFESGSKKCIQV--GGEFYTPSKFEDSG 240
DB 459 RASRKHKDETVDKAPLLPVTCGVGKILHKKRLQOGLIVKCIOTEDGKWFTEPEFIK 518
QY 241 SGKNARSSSGPRPLVRAKGAOGAAGGGEAR--LGOOGSVAPPLAL----- 285
DB 519 ---GHARSKNW--RLSVRC-----GGMPLRLMENGFLPDPPIRYRKRRKRLKSO 566
QY 286 -PSDQLHOKNDECAVCRDGGELICDGCPRAFHILACLSPLREIPSGTWRCSSC-- 340
DB 567 SSVDCM--RNLDECEVCRDGELFCDCDSRVFHEHCHIPPV--EAERTPMNCIFCMA 623
QY 341 ----LQATVOEVQRAEPRPQE 359
DB 624 SPGSQCCQESSEVLERQKCPPE 645

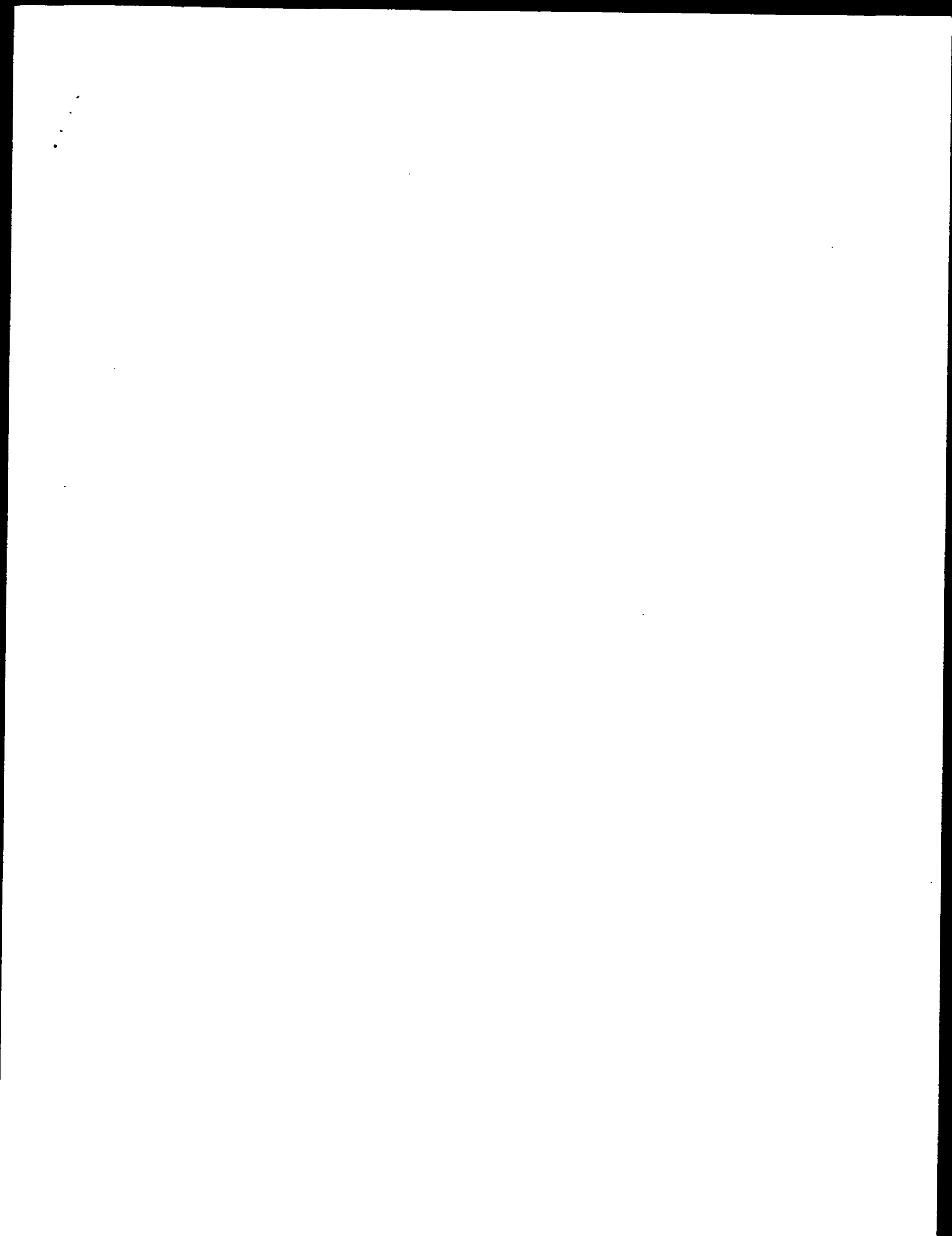
RESULT 15

```

ABG11257
 ID ABG11257 standard; Protein: 771 AA.
 AC
 XX ABG11257;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #11248.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR N-PSDB; NAST5444.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20: SEQ ID No 41616; 103bp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 XX Sequence 771 AA:
 SO
 Query Match 7.4%; Score 213.5; DB 22; Length 771;
 Best Local Similarity 20.7%; Pred. No. 1.3e-07;
 Matches 129; Conservative 54; Mismatches 148; Indels 291; Gaps 26;
 OY 9 RLRLHRTETIAVADSAFLHLADHDVPE---DKFOETLHKEKGGCPQAFHALIS 65
 DB 47 RFRRENKVEIASITRPFPPLGLRDRSFTSEQMYEHFOEA--FRNLVPTVTRVWCYVSE 104
 OY 66 LITODSTALIDFWLVLFKDYNIERYGNLQPIIDLSF-----PKDV 104
 DB 105 L---EKTFQMSHEALFSRINLMAYPDLDNIYRSFQWVCYEHSPLOMNNVNDLEDRPLL 161

OY 105 -----DLSPRK-----GRKP-----AVPKA-----LVP 124
 DB 162 PYGKQENSNNACHEMDLAVQDALSSSARCEPGSSSECEQLALPKAGGDAEDAPSLP 221
 OY 125 -----PRLP-----TRKASE----- 136
 DB 222 VSCKLAIQIDGEESEEMPRLPYDTEFTDLKTPQVTNEGPEPKGLCLPGEGBSDOC 281
 OY 137 -----PARAAPALIPRGAS-----PSQLKA----- 160
 DB 282 SEMCDGEERDASSSLARRGVSSELENHPNNEGESEELASSLLYDVPAGQAYENE 341
 OY 161 -----KPKKPSSAS-----QORLPNG-----I 181
 DB 342 KCSVMCFSEVPSPPEARTESDQAGTMDTVDIANNSTLGPKPKRRKRGHSMRM 401
 OY 182 QTMASAVQRAVAMSSGDV-----PGAR----- 203
 DB 402 RROKNSOQNDNSKADGVVSSSEKRVNVLKDLKIRGRKRGKPGTRFTQSDRAACKRVRS 461
 OY 204 -----GAVEGILIOVFESGSKCICQY-GGEFTYPSKFEDSG 240
 DB 462 RASRKHDEYDFKAPLLPYTCGGYKGLHKKKLQOGLIVKCIQTEDGKMTPTFEIKG 521
 OY 241 SGKNKARSSSGPKPLVRAKGAQGAAPGGGEAR-IGQGSVPAPLAL----- 285
 DB 522 ---GHARSKNM-RLSVRC-----GGWPLRLWLMENGLPDPPIRYRKKRIILKSQNN 569
 OY 286 -PSPQLHKQKNEDECACVARDGELICDGCPRAFHLACLPFLRELPSGTWNCSSC---- 340
 DB 570 SSVDPCLM--RNLIDCEVCGRDGELEFCDCSHVFHEDCHIPPV-EAERTPMWCIFCRMKE 626
 OY 341 ---LOATVQEVOPRAEPRPOE 359
 DB 627 SPGSQCCQCESEVLERQMCPEE 648

Search completed: March 13, 2003, 17:52:52
 Job time: 43 secs



GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:54:02 ; Search time 17 Seconds
(without alignments)
1477.658 Million cell updates/sec

Title: US-09-509-595b-2

Perfect score: 2902

Sequence: 1 MATDALARLRRLRHLRREINAV.....DGLIQAQIAQSMARPAAPFPS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubppaa/PTCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
5: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubppaa/PTCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	168	5.8	1134	9	US-09-836-392-16
2	168	5.8	1237	10	US-09-862-027-78
3	166	5.7	549	10	US-09-764-864-1254
4	165	5.7	167	9	US-10-005-057A-31
5	159.5	5.5	1023	9	US-09-893-519A-14
6	156	5.4	432	10	US-09-764-864-1184
7	155	5.3	1400	10	US-09-764-176-7
8	153	5.3	1207	9	US-10-189-971-20
9	151.5	5.2	2462	9	US-09-819-104A-5
10	149	5.1	601	10	US-09-925-301-844
11	148.5	5.1	360	10	US-09-764-176-3
12	147.5	5.1	538	9	US-09-976-740-43
13	147.5	5.1	538	12	US-10-023-529-43
14	147.5	5.0	538	12	US-10-023-523-43
15	145.5	5.0	1251	9	US-10-189-971-16
16	143	4.9	527	9	US-09-854-133-216
17	143	4.9	527	10	US-09-738-973-216
18	142.5	4.9	510	10	US-09-866-562-56
19	140.5	4.8	1274	9	US-10-020-215-2

20	140	4.8	2507	9	US-09-819-104A-2	Sequence 2, Appli
21	139.5	4.8	1056	9	US-10-161-510-10	Sequence 10, Appl
22	139	4.8	550	9	US-09-976-740-47	Sequence 47, Appl
23	139	4.8	550	12	US-10-023-529-47	Sequence 47, Appl
24	139	4.8	550	12	US-10-023-523-47	Sequence 47, Appl
25	136	4.7	522	9	US-09-764-868-1138	Sequence 1138, Ap
26	135.5	4.7	2441	12	US-10-109-886-8	Sequence 8, Appl
27	135	4.7	524	9	US-09-764-868-761	Sequence 761, App
28	135	4.7	846	10	US-09-858-664A-3	Sequence 3, Appl1
29	134.5	4.6	693	9	US-10-029-217A-4	Sequence 4, Appl1
30	134	4.6	1192	9	US-10-189-971-18	Sequence 18, Appl
31	134	4.6	1342	9	US-10-189-971-24	Sequence 24, Appl
32	134	4.6	1477	9	US-10-189-971-8	Sequence 8, Appl1
33	134	4.6	1512	9	US-10-189-971-10	Sequence 10, Appl
34	134	4.6	1535	9	US-10-189-971-14	Sequence 14, Appl
35	134	4.6	1570	9	US-10-189-971-12	Sequence 12, Appl
36	134	4.6	1593	9	US-10-189-971-4	Sequence 4, Appl1
37	134	4.6	1628	9	US-10-189-971-2	Sequence 2, Appl1
38	133.5	4.6	1002	9	US-09-988-117-3	Sequence 3, Appl1
39	133.5	4.6	1002	10	US-09-812-471-3	Sequence 3, Appl1
40	133.5	4.6	1002	10	US-09-812-633-3	Sequence 3, Appl1
41	133	4.6	1317	10	US-09-963-896-7	Sequence 7, Appl1
42	131.5	4.5	420	10	US-09-764-864-1084	Sequence 1084, Ap
43	131.5	4.5	1690	10	US-09-788-043C-5	Sequence 5, Appl1
44	131	4.5	4019	9	US-09-854-133-425	Sequence 425, App
45	131	4.5	4019	10	US-09-738-973-425	Sequence 425, App

ALIGNMENTS

RESULT 1
US-09-836-392-16
; Sequence 16, Application US/09836392
; Patent No. US20020173458A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Protein Tyrosine Kinase Receptor Polynucleotides, Polypeptides
; FILE REFERENCE: PTO20P1
; CURRENT APPLICATION NUMBER: US/09/836,392
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: PCT/US00/28066
; PRIOR FILING DATE: 2000-10-11
; PRIOR APPLICATION NUMBER: 60/159,542
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: 60/165,914
; PRIOR FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: 60/189,027
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 1134
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-836-392-16
Query Match 5.8%; Score 168; DB 9; Length 1134;
Best Local Similarity 26.4%; Pred. No. 0.01;
Matches 111; Conservative 41; Mismatches 180; Indels 88; Gaps 22;
QY 116 PAVKALVP-----PRLPTKRRKASEEARAAAPALTPPGTASPGSQLAKAPPKRESS 169
DB 767 PLSPLAHTPSPQASPPPLPGHTVGSHTQSPPAKLHSSPPVVRPPKSAEPPRSLK 826
QY 170 ARQQRPLRNGIOITMSSAQAQAVAMSSGV--PGARAVBG-ILIQVFFSGSKKIOY 226
DB 827 RYQSAEKLQ---ASLSADKKGLAKRHSLEVGHPDFKDFGELALHSLASDDETPVEG 883
QY 227 GGEFTYPSKFEEDSGSKNRRSSGPKPLVRANKA-----OGAAPGGGEARLGGQS 278
DB 884 LG---APRGVAVRRLGRQSPSLIGADPLL--PGASRPPVSSKESKESPGCAEACTPPRAT 939

```

QY 279 VPAPLALPSPD--OLHQKNEDECAVCRDGGELICDGCPRAPHLACTSPRLREIPSGTMR 336
D 940 TPGRTLERDVGTCTRHQSVQTE-----DG-----TGMAARAAYAKALIS--PYDHEHG--R 986
QY 337 CSSCLQATVQEVQRAEPPROEPPEVPEPLPGLRSAGEEVGP--PGEPLA--GMD--- 389
D 987 RSSSGEAGTPLVPIVEPARRGAKAV--VPOPLGADSKGQDEPAPLAPSVPEAPRGHERNV 1045
QY 390 -----TTLV-YKHLPAAPSAAPLPGIDSSALHPLVCGPEGQONLAPGACGVCGDGT 441
D 1046 LEVEERTTILSGPRSKRPSPLSPEP--QTPSLAPAKCSAPSSAVTPVPPA--SLIGSGT 1101
QY 442 DVLRCTHCAAAFHMRCHPACTSRPGTGLRCRSCSGDVTAPAV--EGVLAPSPARLAPGP 499
D 1102 -----KPOVGLTSCRPAEAVPPAGLTKKGVSSPAP----PGP 1134

```

RESULT 2

```

US-09-862-027-78
; Sequence 78, Application US/09862027
; Patent No. US20020142428A1
; GENERAL INFORMATION:
; APPLICANT: Hodge, Martin R.
; TITLE OF INVENTION: No. US20020142428A1el Kinases and Uses Thereof
; FILE REFERENCE: 35800/234862
; CURRENT APPLICATION NUMBER: US/09/862,027
; PRIOR APPLICATION NUMBER: 2001-05-21
; PRIOR FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 82
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 1237
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-862-027-78

```

```

Query Match
Best Local Similarity 5.8%; Score 168; DB 10; Length 1237;
Matches 111; Conservative 41; Mismatches 180; Indels 88; Gaps 22;

```

```

QY 116 PAVPALVP-----PPRLPTKKASEEAAAPALTPRGASPGSQLAKPKPKRESS 169
D 870 PLSPLAHPPSPQASPPPLPGHTVSGSHHTQSPFPAKLHSSPPVVRPRPKSAEPPRSLPLK 929
QY 170 AEOQRLPLNGIQTMSASVORAVAMSSGDV--PGARGAVEG--ILIQVFESSGSKKCIQY 226
D 930 RVQSAEKIG---ASLSADKKGALKRKHSLEVGHDFRKKDFHGLALHSLASDGETPVEG 986
QY 227 GGEFTTPSKFEDSGSGKNKARSSSGPKPLVRAKA-----OGAAPGGGEARLGGQGS 278
D 987 LG---APROVAVRRLGRQESPLSLGADPLL--PEGASRPVSSSEKESPGCAEACTPPRAT 1042
QY 279 VPAPLALPSPD--OLHQKNEDECAVCRDGGELICDGCPRAPHLACTSPRLREIPSGTMR 336
D 1043 TPGRTLERDVGTCTRHQSVQTE-----DG-----TGMAARAAYAKALIS--PYDHEHG--R 1089
QY 337 CSSCLQATVQEVQRAEPPROEPPEVPEPLPGLRSAGEEVGP--PGEPLA--GMD--- 389
D 1090 RSSSGEAGTPLVPIVEPARRGAKAV--VPOPLGADSKGQDEPAPLAPSVPEAPRGHERNV 1148
QY 390 -----TTLV-YKHLPAAPSAAPLPGIDSSALHPLVCGPEGQONLAPGACGVCGDGT 441
D 1149 LEVEERTTILSGPRSKRPSPLSPEP--QTPSLAPAKCSAPSSAVTPVPPA--SLIGSGT 1204
QY 442 DVLRCTHCAAAFHMRCHPACTSRPGTGLRCRSCSGDVTAPAV--EGVLAPSPARLAPGP 499
D 1205 -----KPOVGLTSCRPAEAVPPAGLTKKGVSSPAP----PGP 1237

```

RESULT 3
US-09-764-864-1254

```

; Sequence 1254, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; PRIOR APPLICATION DATA REMOVED - consult PAM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1254
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (21)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (290)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (297)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (464)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1254

```

```

Query Match
Best Local Similarity 5.7%; Score 166; DB 10; Length 549;
Matches 61; Conservative 31; Mismatches 93; Indels 60; Gaps 10;

```

```

QY 115 PPAVKALVPPRLPTKKASEEAAAPALTPRGASPGSQLAKPKPKRESSAQR 174
D 334 PPGPTSTANPSP-----GPSALSP--GSSGLSNSTHPVPRPSTSTSGSR 375
QY 175 LPLNGIQTMSASVORAVAMSSGDV--PGARGAV--EGILIQVFESSGSKKCIQY 226
D 376 GSCGSSGRT---AEKTSLSFKSDQYKVKQEPGTEDICFSFGVCKQEKEDGRSCMIS 432
QY 227 GGE-----FTTPSKFEDSGSGKNKAR-----SSSGPKPLVRAKAQAAP 266
D 433 SPESLTPPLSTNLLESELALASLENNVKKXPADMNESCKGSLSLVNGKSPIRSLM 492
QY 267 GGEARLGGQGSVPAPLALPSPDOLHQKNEDECAVCRDGGELICDGCPRAPHLACTSP 326
D 493 -HRAARIGDGN-----NKDDP-----NEDWCACVQNGDILLCEKCPKVFHLCVPT 541
QY 327 LREIP 331
D 542 LLSFP 546

```

RESULT 4

```

US-10-005-057A-31
; Sequence 31, Application US/10005057A
; Patent No. US20020170087A1
; GENERAL INFORMATION:
; APPLICANT: Tao, Yunlin
; APPLICANT: Gordon-Kamm, William J.
; APPLICANT: Shen, Bo
; APPLICANT: Lowe, Keith S.
; APPLICANT: Danilevskaya, Olga
; APPLICANT: Mahajan, Pramod
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Sakai, Hajime
; APPLICANT: Klein, Ted M.
; TITLE OF INVENTION: Transcriptional Regulator Nucleic Acids,
; FILE REFERENCE: 1288
; CURRENT APPLICATION NUMBER: US/10/005,057A

```


; CURRENT FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: 60/251,555
 ; PRIOR FILING DATE: 2000-12-06
 ; NUMBER OF SEQ ID NOS: 41
 ; SOFTWARE: FASTSEQ for Windows Version 3.0
 ; SEQ ID NO 31
 ; LENGTH: 167
 ; TYPE: PRT
 ; ORGANISM: oryza sativa
 ; US-10-005-057A-31

Query Match 5.7%; Score 165; DB 9; Length 167;
 Best Local Similarity 55.8%; Pred. No. 0.002;
 Matches 24; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 298 ECACVRDGEELICDGCRAFHACISPLREIPSGTRCSCG 340
 DB 105 ECVECDLGGNLCCDSCRTYHLECLNPLKRAPPGNMOCPC 147

RESULT 5
 ; US-09-893-519A-14
 ; Sequence 14, Application US/09893519A
 ; Publication No. US20030027243A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ANADYS PHARMACEUTICALS, INC.
 ; APPLICANT: THOMPSON, Craig
 ; APPLICANT: MOORE, Jeffrey
 ; APPLICANT: BURMAN, Ed T.
 ; APPLICANT: BRADLEY, John
 ; APPLICANT: DESILVA, Thamara
 ; APPLICANT: HARRIS, Sandra
 ; APPLICANT: KOMARNITSKY, Svetlana
 ; APPLICANT: MENDILLO, Marc
 ; APPLICANT: MOORE, Daniel
 ; APPLICANT: SANDERSON, Karen
 ; APPLICANT: HAQ, Tariq
 ; APPLICANT: ZHU, Shuhao
 ; APPLICANT: LONG, Fan
 ; APPLICANT: DAVIDOV, Eugene
 ; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
 ; FILE REFERENCE: 0342/1G548-US2
 ; CURRENT APPLICATION NUMBER: US/09/893,519A
 ; CURRENT FILING DATE: 2001-06-28
 ; PRIOR APPLICATION NUMBER: US 60/215,164
 ; PRIOR FILING DATE: 2000-06-29
 ; PRIOR APPLICATION NUMBER: US 60/224,457
 ; PRIOR FILING DATE: 2000-08-10
 ; NUMBER OF SEQ ID NOS: 146
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 14
 ; LENGTH: 1023
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Corresponds to SEQ ID NO: 87
 ; PUBLICATION INFORMATION:
 ; DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
 ; DATABASE ENTRY DATE: 1997-06-25
 ; RELEVANT RESIDUES: (1)..(1023)
 ; US-09-893-519A-14

Query Match 5.5%; Score 159.5; DB 9; Length 1023;
 Best Local Similarity 23.2%; Pred. No. 0.03;
 Matches 109; Conservative 26; Mismatches 163; Indels 171; Gaps 22;

QY 126 PRLLTKRKAEEAARAAPAL-----TPRGT--ASPSQAKAPPKPPKSSAEGQRUP 176
 DB 45 PRTP-----EVRAAAAGALGNHVVSSPAGACAGPRAAPGAPGAAPEPPPPACRAP 97
 QY 177 LNCIGTWSASVQRAVAMSSGDVPCA---RGAVGILIQVVESSG-----SKKCIQVG 227

DB 98 GGGGQRRGPPPPRRPLVPAGPAPPAKLRPPPEG-----SAGACAPVPAAMAAYAG 149
 QY 228 GEFTYPSKFEDSGSGKNNKARSSGPKPLVRAKGAAGAPGGGEALGGQGSV----- 279
 DB 150 PEPAAPAKPAGPAAALAAAGPGGPGP-----GPPGPGPGKPPAGPAGAGQTNGSAA 201
 QY 280 -----PA---PLALPSDPOLHOKNDECAVCRDGGELICDCCDPCPAF 318
 DB 202 LNSHHAAPAVSLVNNGPALLLPKPAAP-----GTVI-----QTRP 240
 QY 319 HLACLSPPLREIPSGTRCSCGLOATVOEVPRAEPPRQEPVETPLPGLRSAGEEVR 378
 DB 241 FVGAAPAPPAAPPS-----PPAA--PAPAAAPAAPPP--PAPATLA 279
 QY 379 GPPEGLAGMDTTLVYKHLPAAPSAAPLPGDSSALHLLCGVPGQQLNAPGARGCGV 438
 DB 280 RPPGHP-AG-----PPTAAP-----AVPPAAANQGSAGAAPAPAPAAAG 319
 QY 439 DGTDLRCHCAAFHMRCHPAGTS--RPTGLRCRSCSGDVTAPAVEGVLPASPARL-- 495
 DB 320 -----PAGVSGPGPGA-----AAAPAPGVRAESPKRVVQ 350
 QY 496 APGPAKDDTASHEPALHRDLLESLSEHTPDGILQWAIQSMARRAPAP 544
 DB 351 AAPPAOTLAAAGDA-----STAASWVIGPTMGALPSPAAPPAP 392

RESULT 6
 ; US-09-764-864-1184
 ; Sequence 1184, Application US/09764864
 ; Patent No. US20020132753A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PT23
 ; CURRENT APPLICATION NUMBER: US/09/764,864
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 1792
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1184
 ; LENGTH: 432
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (31)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (35)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (38)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (40)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (44)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; NAME/KEY: SITE
 ; LOCATION: (102)
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 ; US-09-764-864-1184

Query Match 5.4%; Score 156; DB 10; Length 432;
 Best Local Similarity 22.3%; Pred. No. 0.019;
 Matches 65; Conservative 39; Mismatches 100; Indels 88; Gaps 12;

QY 111 GKRRPAPVAPVPPPLPTKRKAEEAARAAPALTPRGTASPSQAKAPPKPPKSSA 170
 DB 101 RKKKPDSPVPLBEPENKQEKTEKEEKTNGVTRLRSRPR-----ISRPTAVAEIRDKA 155

```

QY 171 EQORLPLGNG---IQTMSAVORA-----VAMSSGDVPGARAGVEGILIQVFESE----- 218
DQ 156 DKR---GEGEDEVEEESTALOKTKKELKSEKDTNKK-----VSKVPPKQKVRW 204
QY 219 -GSKKCIQVGEFTYPPSKFEDSGSKNKARSSSGPKPLVRAGAOGAAGGEGARLGOOG 277
DQ 205 TGRS---TRGRWKYSNDESESGSEKSSAASEEKESE-----EAILADDD 250
QY 278 SVAPRLALPSPOLHOKNEDECAVCRDGGELICDGCPRAFHLACLSPPLREIPSGTWRC 337
DQ 251 EPCKKGLPNHPEL-----ILLCDSCDSGYHTACTLRPLMIIPDGEWFC 294
QY 338 SSC-----LQATVQEV-----QPRAEPR-----POEP 360
DQ 295 PRCOKHLCEKLEBQLODLDVALKKKERARERKRLVYVGISIENTIIPPOEP 346

RESULT 7
US-09-764-176-7
; Sequence 7, Application US/09764176
; Patent No. US20020127553A1
; GENERAL INFORMATION:
; APPLICANT: NOTEBORN, Mathieu Hubertus Maria
; APPLICANT: DANEN-VAN OORSCHOT, Astrid Adriana Anna Maria
; APPLICANT: ROHN, Jennifer Leigh
; APPLICANT: WEISS, Bertram
; TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
; FILE REFERENCE: 472505
; CURRENT APPLICATION NUMBER: US/09/764, 176
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1400
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Description of Sequence: Amino acid sequence deduced from the nuc
; OTHER INFORMATION: Ietc acid sequence of AAP-
US-09-764-176-7

Query Match 5.3%; Score 155; DB 10; Length 1400;
Best Local Similarity 22.4%; Pred. No. 0.078;
Matches 65; Conservative 38; Mismatches 99; Indels 88; Gaps 12;

QY 113 KRPRNVPKALVPPRLPTKKAASEBARAALTPRGTAAPGSQLAKPKPKPSSAEQ 172
DQ 703 KRPSPPKVLPEPNKQEKTEKEEKTNGRTLRSPR-----ISRPTAKVAEIRDQADK 757
QY 173 QRLPLGNG---IQTMSAVORA-----VAMSSGDVPGARAGVEGILIQVFESE-----G 219
DQ 758 KR---GEGEDEVEEESTALOKTKKELKSEKDTNKK-----VSKVPPKQKVRW 806
QY 220 SKKCIQVGEFTYPPSKFEDSGSKNKARSSSGPKPLVRAGAOGAAGGEGARLGOOGSV 279
DQ 807 SR---TRGRWKYSNDESESGSEKSSAASEEKESE-----EAILADDD 852
QY 280 PAPRLALPSPOLHOKNEDECAVCRDGGELICDGCPRAFHLACLSPPLREIPSGTWRCSS 339
DQ 853 CKKGLPNHPEL-----ILLCDSCDSGYHTACTLRPLMIIPDGEWFC 896
QY 340 C-----LQATVQEV-----QPRAEPR-----POEP 360
DQ 897 COKHLCEKLEBQLODLDVALKKKERARERKRLVYVGISIENTIIPPOEP 946

RESULT 8
US-10-189-971-20
; Sequence 20, Application US/10189971
; Publication No. US20030028907A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; TITLE OF INVENTION: No. US20030028907A1el Human Kieelin-like Proteins and Polynuc
; FILE REFERENCE: LEX-0360-USA
; CURRENT APPLICATION NUMBER: US/10/189, 971
; CURRENT FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/302,949
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: US 60/315,634
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 1207
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-189-971-20

Query Match 5.3%; Score 153; DB 9; Length 1207;
Best Local Similarity 22.7%; Pred. No. 0.088;
Matches 99; Conservative 44; Mismatches 164; Indels 130; Gaps 22;

QY 121 ALVPPRLPTKKAASEBARAALTPRGTAAPGSQLAKPKPKPSSAEQORLPLGNG 180
DQ 7 SLCPARQCP---ADQARSLGSAALSAMAVSTRDTSIRARRPDSKSGASVNAAP--- 58
QY 181 IQTMSAVORAVAMSSGDV-----GARGAV-----EGILIQVFESESKKCIQVYG 228
DQ 59 -ARLARSVPKRSKASQSPVPCPLLAASSAOGACCPSCDSTYHSQVYANG----- 107
QY 229 EFTYPPSKFEDSGSKNKARSSSG-----PKPLVRAGAOG--AAPGEGARLGOOG 277
DQ 108 ---QNTDADSPCHACHQODGTVCSLVDCPRTTCARQSGPCCPCPCDILBEV 162
QY 278 SVAPRLALPSPOLHOKNEDECAVCRDGGELICDGC-----CPRAFHLACLSP--- 325
DQ 163 EV-----DSEFSHPRPDQECR-----COEGHANQCPRPCCRA---PCANPLPRT 205
QY 326 -----PLREIPSGT-----WRCSSCLQATVQEVQPR-----AEEP----- 355
DQ 206 CCPNDCSGCAFGEKYEPSADPRHPSDPCHLCKLSGNVOCLARRCVLPRLCPREVLLPGE 265
QY 356 -RPQEP---PYETPLPRLKRSAGEEVGRPPGEPYLAG---MDTLYYKHLRPPS---AAP 405
DQ 266 CCRQCPAARAPAGCPRPGAANAHNDEYFSPRQDPCRCCLDGSVSCQRLPCRPACANP 325
QY 406 LRGDSSALHPLLYGPE---GOONLAPCARCG--VGGGTVDYLKCTHCAAFNRCHRP 460
DQ 326 RQGPCPCSCDGLYQGEKESASGERPPYTAACHLCLWEGSVSCPEKACAPAL---CPFP 382
QY 461 AGTSRPGTGLRCRCSG 477
DQ 383 A-----RGDCCPDCDG 393

RESULT 9
US-09-819-104A-5
; Sequence 5, Application US/09819104A
; Publication No. US20030027137A1
; GENERAL INFORMATION:
; APPLICANT: Chen, J. Don
; TITLE OF INVENTION: NOVEL NUCLEAR RECEPTOR COREPRESSOR MOLECULES
; FILE REFERENCE: UMG-030
; CURRENT APPLICATION NUMBER: US/09/819, 104A
; CURRENT FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: 60/193,138
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5

```

LENGTH: 2462
 TYPE: PRT
 ORGANISM: Mus musculus
 US-09-819-104A-5

Query Match 5.28; Score 151.5; DB 9; Length 2462;
 Best Local Similarity 22.98; Pred. No. 0.23;
 Matches 121; Conservative 42; Mismatches 186; Indels 179; Gaps 24;

98 DSFKVDLSQPRKGRK-----PAYKALVPPRLPTKRASBARAAALTPRGT 151
 746 ESVPSPSEAKKDTGPKRTGEALPAATQPPVPEEP-----AAAPAEBSVPD 795
 152 AS-PSQLAKAPPKPPESSA--EQORLPLGNIGQTMASVQRAVA-----MSG 197
 796 ASGPSPSPSPAPATVQKDEQEAAPAPQTEDEAKQSEAEIDVQKPEEPASE 855
 198 DVP-----GARG--AVEGILIQV-FESGSKKCIQVGEFFYPSKFE 237
 856 EPPSPVSKDHKEETEEDKAKGTETVSEAPLKVEASKAAYTKGS--SSGATQ 912
 238 DSGS-----GKNKAR-----SSGPKPLVRAKAGQAA-- 265
 913 DSDSSATCSADEVDEPEGDKGRLLSPRSLTPAGDPRASTSPQKPLDKOLKORAAAI 972
 266 -----PGGEARLQOGSVAPARLALPSDPLHOKNEDECAVCRDGGELICDGP 315
 973 PPIATVKEHP-----REDTVP-PKRVPPVPPPTQHLOREGEVDSQSG-----GSP 1017
 316 RAFLACISPLRLPSTGTMKSSCLQATVOEVOPEAEPPPEPVEP----- 364
 1018 RGRSRSPVPAEKAEKAPFAPFAPTEGOSYRLSPHAGHRLPSHPREVIKTSTRADPLF 1077
 365 -----PLPGLSAGEVEVGPPEPLAGMDTLVYKHLRPAAPALPGDSSALHPL 417
 1078 SYTPPGHPLPLGLHDSARV-----LPPPLPSNP-PLLISSAKHP- 1116
 418 LCVPEGOQLNLPARGCVCGDGTDV-LRCHCAAAFMHRCHEPAGTSRPTGLRCRSCS 476
 1117 ---GVLEHQ-----LGAISQMSVQLRVPHSEHA-----KPMGPLMELPLA----- 1155
 477 GDVTPAVEGVLPAPARLAPGAKDPTASHEPALHDDLESLEHT 524
 1156 ---VDPKRLGTALAPPVPEASP-----RASQYPGCRFPQLRLYHPR 1195

RESULT 10
 US-09-925-301-844
 Sequence 844, Application US/09925301
 Patent No. US20020052308A1
 GENERAL INFORMATION:
 APPLICANT: Rosen et al.
 TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 FILE REFERENCE: PA106
 CURRENT APPLICATION NUMBER: US/09/925,301
 CURRENT FILING DATE: 2001-08-10
 PRIOR APPLICATION NUMBER: PCT/US00/05882
 PRIOR FILING DATE: 2000-03-08
 PRIOR APPLICATION NUMBER: 60/124,270
 PRIOR FILING DATE: 1999-03-12
 NUMBER OF SEQ ID NOS: 1694
 SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 844
 LENGTH: 601
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: SITE
 LOCATION: (36)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (64)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE
 LOCATION: (103)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (106)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (152)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (358)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 NAME/KEY: SITE
 LOCATION: (383)
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
 US-09-925-301-844

Query Match 5.18; Score 149; DB 10; Length 601;
 Best Local Similarity 24.18; Pred. No. 0.073;
 Matches 121; Conservative 33; Mismatches 175; Indels 174; Gaps 27;

92 RLQPIIDSPFKVDV-LSQ-----PRKGRKPAVAKALVPPRLPTKR-- 133
 82 RLQVVIHSNPAQIQOALALSLXROXSVTAPGHRHKKRPPQAVQG--PSPRPRTYEPQ 139
 134 -----ASEEARAAAPAL-----TPRGTA--SP-GSOLKAKPKPKPSSAEQRLPIGN 179
 140 RVNSGLSSDHFKEPFGVMVGVGCTPRDSAGVSFPFRKRRRPRPKRLDLESILP--- 196
 180 GIOTMSASVQRAVAMSSGDVPGARGAVEGILIQVFESGSKKCIQVGEFFYTP--SKF 236
 197 -----PHSSGFL-GSK--PEG-----PGQAESNDTGTETALPHIMNRL 232
 237 EDGSGCKNKARSSGP-----KPLVRAKAGQAAAPGGEALQOGSVAPARLALPSDQL 291
 233 HTATSRKSYRSPMEPMEPLSPEDVAGTEMSSGSDVLSGDSQVSSGCSQSSP-- 290
 292 HOKNEDECAVCRDGGELICDGCPRAFHLACTSPRLPSTGTMKSSCLQATVOEVOPE 351
 291 -----DGLKGAAG-----PPKR--PEG-----SSPLNAVCEGEPG 321
 352 AEPRPQEP-----PVETPLPG-----LRSAGEEYGP-----PGEP-- 384
 322 SEPPRRRPPAPHDQKRELPEGLPPGPICTERSQXTDRCTE--GPIRPSHRPQVQF 380
 385 -----LAGMDTLVYKHLPA-----PSAAPLPGDSSALHPLLCV 420
 381 GTXDKDSDLRLVVGDSLKAKEELTASVTEALPVSRDWELPSAASAPQSKNLDSCGCV 440
 421 GPE---GQNLAPGARGCVCGDGTDVLRCHCAAAFMHRCHEPAGTSRPTGLRCRSCSG 477
 441 -PEPSSGQRLYEVFVFGSAGPSSQI-----SGGAMSQQLHPNSGGFRPGT----- 486
 478 DVTAPAVEGVLPAPARLAPGPA 500
 487 -----PSLHPRYSQPLVLPQGA 504

RESULT 11
 US-09-764-176-3
 Sequence 3, Application US/09764176
 Patent No. US20020127553A1
 GENERAL INFORMATION:
 APPLICANT: NOTEBOEN, Mathieu Hubertus Maria
 APPLICANT: DANEN-VAN OORSCHOT, Astrid Adriana Anna Maria
 APPLICANT: ROHN, Jennifer Leigh
 APPLICANT: WEISS, Bertram
 TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
 FILE REFERENCE: 4725US
 CURRENT APPLICATION NUMBER: US/09/764,176
 CURRENT FILING DATE: 2001-01-17
 NUMBER OF SEQ ID NOS: 18
 SOFTWARE: Patentin version 3.0

SEQ ID NO 3
LENGTH: 386
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Description of Sequence: amino acid sequence of the analyzed regi
OTHER INFORMATION: on of the Apoptin-associating clone AAP-2-II
NAME/KEY: misc_feature
LOCATION: (1)-(3)
OTHER INFORMATION: Description of Sequence: the three C-terminal amino acids (His-G)
US-09-764-176-3

Query Match 5.1%; Score 148.5; DB 10; Length 386;
Best Local Similarity 23.2%; Pred. No. 0.049;

Matches 56; Conservative 32; Mismatches 94; Indels 59; Gaps 9;

QY 113 RKPAPVAKALVPPRLPTKRKASEARAAAPALTPRGTAASPSQLKAKPKPPKSSARO 172
DB 154 KKPDPSPKVLPEPKKQEKTEKEETKNTVGRRLRSR-----ISRPYAKVAEIRQKADK 208
QY 173 QRLPLGNG---IQTMASVORA-----VAMSGDVPGARGAIVEGILLIOVFESG-----G 219
DB 209 KR--GGEDEVEEESTALQTDKREILKSEKDTNSK-----VSKVKPKGVWMTG 257
QY 220 SKKCIQVGEFFTPSKFEDSGSKNKARSSGPKPLVRAKGAAGAPGGEARLQOQSV 279
DB 258 SR-----TRGMWYSSNDSESGSGSEKSSAAEEEEKSE-----EAILADDEP 303
QY 280 PALPALPSDQLQHKNDCEAVCRDGGELICDCCPRAFHILACTSPPLREIPSGTWRCSS 339
DB 304 CKKCGJPNHPEL-----ILLDCSDSGYHACLRPLMLITPDGEMFCPP 347
QY 340 C 340
DB 348 C 348

RESULT 12

US-09-976-740-43
Sequence 43, Application US/09976740
Publication No. US20020194633A1
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.
APPLICANT: Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/09/976,740
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 09/616,289
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 43
LENGTH: 538
TYPE: PRT
ORGANISM: Homo sapiens
US-09-976-740-43

Query Match 5.1%; Score 147.5; DB 9; Length 538;
Best Local Similarity 22.6%; Pred. No. 0.08;
Matches 106; Conservative 47; Mismatches 136; Indels 181; Gaps 26;

QY 56 POAFHALLSWLLTQDSTAILDFWVLFK-DVNLERYGRLOPLDSEFPKDVDSQPRKGRK 114
DB 66 PERTRAELEKLIQ--RAVL---KSTYKSTISYRAARVOP-----PRRCAT 107
QY 115 PPAVPAK-----LVPPRLPTKRKASEEA-RAAPALTP--RGTAAP 154
DB 108 PPAPPAAPRGAPAAAAAAPPPTAPPPAPVAAAAAPARAAPAAAAATAPSPGPAP 167
QY 155 GSOLK-----AKPKKPESSAQOR-- 174
DB 168 GPRAPQRAAPLAAPPAPAPVAPAGPRAPAPVAPAPVAPAPVAPAPVAPAPVAP 227
QY 175 -----LPLNGIQTMSASVORAAM-----SSGDPG--ARGAVEGILLIOVFESG 218
DB 228 PPQPPQPPPEGAVARGAA--RPVSLDEVVYTLGSGGAGRLRGVQGLLEEAARG 285
QY 219 GSKKCIQVGEFFTPSKFEDSGSKNKARSSGPKPLVRAKGAAGAPG-----GGEARLG 274
DB 286 -----RLERTRLG--ALALPRGDRP--GRAPASARSRSKRGEEKVL 326
QY 275 QOQSVAPALPSDQLQHKNDCEAVCRDGGELICDCCPRAFHILACTSPPLREIPSGT 334
DB 327 EK-----EEEDDDDEDEEDVSGSEV-----PESDRPAG-- 358
QY 335 WRCSCLQATVQDEVPRAE--PRQEPVETPLPGLRSAGEEVRGP--PGE-----PLA 386
DB 359 -----AQHQLNGEGRPOSARKERVKEMTPCGHQODEGRGAPSGTRQVPSMA 408
QY 387 GMD-----TLVY-----KHLPAAPSAAPLPGLDSSALHPLVGVPEGOQ 426
DB 409 ANKKEGTASVATGPDSPSPVPLPGLKALPGADG---PFGC--PPGRK 453

RESULT 13

US-10-023-529-43
Sequence 43, Application US/10023529
Patent No. US20020129388A1
GENERAL INFORMATION:
APPLICANT: Lees, Ann M.
APPLICANT: Lees, Robert S.
APPLICANT: Law, Simon W.
APPLICANT: Arjona, Anibal A.
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
FILE REFERENCE: 10797-004001
CURRENT APPLICATION NUMBER: US/10/023,529
PRIOR FILING DATE: 2001-12-17
PRIOR APPLICATION NUMBER: 09/616,289
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: US 09/517,849
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: US 08/979,608
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 60/031,930
PRIOR FILING DATE: 1996-11-27
PRIOR APPLICATION NUMBER: US 60/048,547
PRIOR FILING DATE: 1997-06-03
NUMBER OF SEQ ID NOS: 53
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 43
LENGTH: 538
TYPE: PRT
ORGANISM: Homo sapiens
US-10-023-529-43

Query Match 5.1%; Score 147.5; DB 12; Length 538;
Best Local Similarity 22.6%; Pred. No. 0.08;
Matches 106; Conservative 47; Mismatches 136; Indels 181; Gaps 26;
QY 56 POAFHALLSWLLTQDSTAILDFWVLFK-DVNLERYGRLOPLDSEFPKDVDSQPRKGRK 114

```

Db      66 PERTRAELEKLIQ--RAVL--RVSYKGSISYRNAARVQ-----PARGAT 107
Qy      115 PRAVPA-----LVPPRLPYTKKASEEA-RAAAPALTP--RGTASP 154
Db      108 PPRAPRPRGAPAAAAAAAPPTPPAPPRPAPVAAAAAPARAPAAAAATAPPSPGAP 167
Qy      155 GSOLK-----AKPKPPESSAEOQ-- 174
Db      168 GPRAQRAAPLAAPPAPAPAPVAPAGRRAPRAAPVAAAREPRLPPPGPPAPAQOQOP 227
Qy      175 -----LPLNGIQTMSASVQRAVAM-----SSGVPFG--AGCAVGGILIQVFE 218
Db      228 PPOPOPPEGGAVRAGAA--RPVSLREVVYVYLGSGGAGGRLTGRGVGLLEEEAARG 285
Qy      219 GSKKCIQVGEFTTPSKFEEDSGSSKNKARSSGPRPLVRKAQAQAAG--GGAARLG 274
Db      286 -----RLERTRLG--ALALPRGDR-GRAPMAAARPSRSKRGGEVYL 326
Qy      275 QQGSVAPRLALPBDPOLHQKNEDCAVOCROGELLCCGCRPAFLIACLSPLRELPST 334
Db      327 EK-----DEEDDEDDDEDDVSEKSEV-----PESDRPAG- 358
Qy      335 WRGSSCLQATVQVQRAAE--PRQEPPEVETPLRPGLSAGEEVGR-PGE-----PLA 386
Db      359 -----AQHHQLNGERGPQSAKERYKWTTPCGPHGOODEGRGPAAGSGTROYFSMA 408
Qy      387 GMD-----TTLVY-----KHLRPPSAARPLGLDSALHPLLCVGPBGOQ 426
Db      409 AMNKEGTISVATGPDPSPSPVPLRPGKRALGADT---PFGC--PPGK 453

```

```

01 RESULT 14
02 US-10-023-523-43
03 ; Sequence 43, Application US/10023523
04 ; Patent No. US20020152485A1
05 ; GENERAL INFORMATION:
06 ; APPLICANT: Lees, Ann M.
07 ; APPLICANT: Lees, Robert S.
08 ; APPLICANT: Law, Simon W.
09 ; APPLICANT: Arjona, Anibal A.
10 ; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
11 ; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
12 ; TITLE OF INVENTION: ATHEROSCLEROSIS
13 ; FILE REFERENCE: 10797-004001
14 ; CURRENT APPLICATION NUMBER: US/10/023, 523
15 ; CURRENT FILING DATE: 2001-12-17
16 ; PRIOR APPLICATION NUMBER: US/09/616, 289
17 ; PRIOR FILING DATE: 2000-07-14
18 ; PRIOR APPLICATION NUMBER: US 09/517, 849
19 ; PRIOR FILING DATE: 2000-03-02
20 ; PRIOR APPLICATION NUMBER: US 08/979, 608
21 ; PRIOR FILING DATE: 1997-11-26
22 ; PRIOR APPLICATION NUMBER: US 60/031,930
23 ; PRIOR FILING DATE: 1996-11-27
24 ; PRIOR APPLICATION NUMBER: US 60/048,547
25 ; PRIOR FILING DATE: 1997-06-03
26 ; NUMBER OF SEQ ID NOS: 33
27 ; SOFTWARE: FastSeq for Windows Version 4.0
28 ; SEQ ID NO 43
29 ; LENGTH: 538
30 ; TYPE: PRT
31 ; ORGANISM: Homo sapiens
32 US-10-023-523-43
33
34 Query Match 5.18; Score 147.5; DB 12; Length 538;
35 Best Local Similarity 22.68; Pred. No. 0.08;
36 Matches 106; Conservative 47; Mismatches 136; Indels 181; Gaps
37
38 04 56 PAFHALLSWLQDPAIIIDFWVLEK-DYNLEHYGRIQPLIDSPKXVDLSQPKGRK 114
39 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
40 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
41 Db 66 PETRAALELEKIQD--RAVL--RVSTKGSISYRAARQDP-----PRGAT 107
42 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
43 115 PPAVPA-----LVPPRLPTKRASEEA-RAAAPALTP--RGTASP 154

```

Dd	108	PPAPRABRGARAAAAAAAAAPPPTAPRPPPPRPAVAAAAARARARAAAAAATAAPSPBGAOP	167
Qy	155	GSOLK-----AKPKKRESSAEOR--	174
Dd	168	GPRAGRAAPPLAAPRPAAPAAVAVAPRAGRARRAPRPAVAAAREPPLPRPPGPRAPRQQOQR	227
Qy	175	-----LPJLNGIOTMSASVORAVAM-----SSGDVPG--ARGAVEGILIOQFESG	218
Dd	228	PPQOPRPEEGAGAVRAGMA--RPVSLBEVYRVLGSGAGAGRLRGHVQGLLEBEAARG	285
Qy	219	GSKKCIQVGGEFYTPSKREDSSGSKNKARSSGKPLRYAKGAOGAAG--GGEARLG	274
Dd	286	-----RLERTRLG--ALALPRGDRP--GRAPRASARPSRSKRGGEERYL	336
Qy	275	QOGSVAPRALPSPDQLHOKNEDECAVCRODGELICDGCGRAPHLACLSPPLREIPSGT	334
Dd	327	EK-----EEDDEDEDEDEDDVSESSEV-----PESDRPAG-	358
Qy	335	WRCSCLQATVOEVPRAEE--PROEPEVPERPLRPLGRLSAGEEYRGF--PGE-----PLA	386
Dd	359	-----AOHHQJLNGERGPQAKERKVKWPTCCGPHQGDGERGAGPASPSTQROVESMA	408
Qy	387	GMD-----TTLVY-----KHLPAPSAAPRLGGLSSALHPLLYCGPRGQO	426
Dd	409	AMNKEGTAAYATGPDSPPVPLPRPKALRGAGGT--PRGC--PRGR	453

```

RESULT 15
US-10-189-971-16
; Sequence 16, Application US/10189971
; Publication No. US20030028907A1
GENERAL INFORMATION:
APPLICANT: Walke, D. Wade
APPLICANT: Scoville, John
APPLICANT: Turner, C. Alexander Jr.
TITLE OF INVENTION: No. US20030028907A1 Human Kielln-like Proteins and Polynucle
TITLE OF INVENTION: Same
FILE REFERENCE: LEX-0360-US-A
CURRENT APPLICATION NUMBER: US/10/189,971
CURRENT FILING DATE: 2002-07-03
PRIOR APPLICATION NUMBER: US 60/302,949
PRIOR FILING DATE: 2001-07-03
PRIOR APPLICATION NUMBER: US 60/315,634
PRIOR FILING DATE: 2001-08-29
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 1251
TYPE: PRF
ORGANISM: homo sapiens
US-10-189-971-16

Query Match          5.0%; Score 145.5; DB 9; Length 1251;
Match Local Similarity 21.7%; Pred.No.0.26;
Matches 97; Conservative 44; Mismatches 166; Indels 139; Gaps 21;

QY      121 ALVPPRLPRKRKASEARAAPAALTTPRGTA SPGSQLKAKPKKPSSAEQRPLGLNG 180
       7 SLCPARQCP-----ADTQRSLSGSALASAMAVSTRDTSIRARRPSDSKSAGSVSNAP----- 58
DB      181 IQTMSASVOARVAAMSSSGVPFCARGAVEGLI-----QQVFESGG 219
       59 -ARTLARSPIVSRSRASAQSPVPCLPTLAASSAQTLTSLPHPGACPCSDCTYYHSQVYANG- 116
QY      220 SKRCIQGGEFYTPSKFEEDSGSGKNRKARSSSG-----PKLVRAKGAGC-AAFGG 268
       117 -----QNFTDADSPCHACHQCDDGTVCSTLYDCPRTTCARPQGSGPQCPCR 162
DB      269 GEARLGQGVSPATLALPDPQLHOKNEDECAVCRGGELICDG-----CPRAFHL 320
       163 PCDILEEEVV-----DSESFSHPRDCCBCR-----COEGHANCPRPBPRA--- 205

```

Thu Mar 13 17:59:50 2003

us-09-509-595b-2.rapb

Page 8

```

0Y 321 AGLSP-----PLERPSGT-----WRSSCLQATVQVEOPR-----A 352
Db 206 PCAHNLPETCCPNDSCGAFGKGYEPGAFPHPSDPCRLCRLSGNVQCLARKCPVLP 255
0Y 353 EEP-----RPEP-----PVETPLPGLSAGEVNGPGEPLAG---MDTLVYKHLR 398
Db 266 PEVLLPGECCPCCPAPAPAPAGCPRPAHAHROEYFSPGDDCRLCLDGSVSQCARP 325
0Y 399 APPS--AAPLPGLDSSALHPLLVGPR--GQONLAPGARCG--VSGDGDVLVRLTHCA 451
Db 326 CPRAPCAHPRGCPCCPCDCLCYGKFEFASGEBRPSPTAACHLCLCMESGVSEEPACAP 385
0Y 452 AFHWRCHPAPGTSRPGTGLCRSG 477
Db 386 AL---CPRPA-----RGDCCPDDG 402

```

Search completed: March 13, 2003, 17:58:32
Job time : 23 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:52:12 ; Search time 16 seconds

(without alignments)
1002.218 Million cell updates/sec

Title: US-09-509-595B-2

Perfect score: 2902

Sequence: 1 MARDALRLRLRLHREIAV.....DGIQWAIQSMARPAPPS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA.*

2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*

3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*

4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*

5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*

6: /cgn2_6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241	8.3	1912	4 US-08-913-832A-2	Sequence 2, Appl
2	241	8.3	1912	4 US-09-249-161A-1	Sequence 2, Appl
3	236.5	8.1	351	5 PCT-US91-06418-1	Sequence 1, Appl
4	213.5	7.4	753	4 US-08-942-686-2	Sequence 2, Appl
5	158	5.4	1593	4 US-08-628-829-4	Sequence 4, Appl
6	154	5.3	2972	4 US-09-579-181-2	Sequence 2, Appl
7	154	5.3	3118	4 US-09-579-181-1	Sequence 1, Appl
8	151.5	5.2	2289	3 US-09-051-019-2	Sequence 2, Appl
9	149	5.1	1384	4 US-08-976-235-11	Sequence 11, Appl
10	147	5.1	1093	3 US-08-545-860D-55	Sequence 55, Appl
11	147	5.1	1093	5 PCT-US94-04496-55	Sequence 55, Appl
12	144.5	5.0	902	1 US-08-396-479B-6	Sequence 6, Appl
13	144.5	5.0	902	1 US-08-818-823-6	Sequence 6, Appl
14	143	4.9	527	4 US-09-370-838-216	Sequence 216, App
15	142.5	4.9	510	4 US-08-246-489-2	Sequence 2, Appl
16	140.5	4.8	1274	4 US-09-095-443-2	Sequence 2, Appl
17	140	4.8	830	4 US-09-562-737-38	Sequence 38, Appl
18	139	4.8	1187	3 US-08-320-559-28	Sequence 28, Appl
19	139	4.8	1187	3 US-08-545-860D-28	Sequence 28, Appl
20	139	4.8	1187	5 PCT-US94-04496-28	Sequence 28, Appl
21	139	4.8	1210	1 US-08-320-559-26	Sequence 26, Appl
22	139	4.8	1210	3 US-08-343-860D-26	Sequence 26, Appl
23	139	4.8	1210	5 PCT-US94-04496-26	Sequence 26, Appl
24	138.5	4.8	1065	1 US-08-642-255-72	Sequence 72, Appl
25	137.5	4.7	1290	4 US-09-150-460B-6	Sequence 6, Appl
26	137	4.7	633	1 US-08-642-255-73	Sequence 73, Appl
27	136.5	4.7	830	4 US-09-562-737-33	Sequence 33, Appl

28	136.5	4.7	882	4 US-09-413-814-78	Sequence 78, Appl
29	135.5	4.7	335	2 US-08-405-175A-6	Sequence 6, Appl
30	135.5	4.7	2441	1 US-08-194-468-2	Sequence 2, Appl
31	135.5	4.7	2441	3 US-08-961-739-2	Sequence 2, Appl
32	135.5	4.7	2441	4 US-09-514-247A-8	Sequence 8, Appl
33	134	4.6	659	4 US-09-562-737-14	Sequence 14, Appl
34	134	4.6	992	1 US-08-127-499A-1	Sequence 1, Appl
35	134	4.6	992	1 US-08-482-847-1	Sequence 1, Appl
36	134	4.6	1185	4 US-09-041-886-23	Sequence 23, Appl
37	133	4.6	1317	3 US-09-083-521-7	Sequence 7, Appl
38	132	4.5	1004	4 US-08-916-352-2	Sequence 2, Appl
39	132	4.5	1050	4 US-09-428-711A-16	Sequence 16, Appl
40	130.5	4.5	1060	3 US-08-931-820-3	Sequence 3, Appl
41	130.5	4.5	1418	3 US-08-963-825-20	Sequence 20, Appl
42	130.5	4.5	1418	4 US-09-010-999-1	Sequence 1, Appl
43	130.5	4.5	1418	4 US-09-500-811-20	Sequence 20, Appl
44	130.5	4.5	1418	4 US-09-570-573-20	Sequence 20, Appl
45	130.5	4.5	1418	4 US-09-548-608-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1	US-08-913-832A-2	US-08-913-832A-2
Sequence 2, Application	US/08913832A	
Patent No. 6329517		
GENERAL INFORMATION:		
APPLICANT: Seelig, Hans Peter		
APPLICANT: Kenz, Manfred		
TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN		
FILE REFERENCE: 8484-0030-999		
CURRENT APPLICATION NUMBER: US/08/913,832A		
CURRENT FILING DATE: 1998-01-12		
PRIOR APPLICATION NUMBER: PCT/DE96/00444		
PRIOR FILING DATE: 1996-03-08		
NUMBER OF SEQ ID NOS: 2		
SOFTWARE: FastSeq for Windows Version 4.0		
SEQ ID NO 2		
LENGTH: 1912		
TYPE: PRT		
ORGANISM: Homo sapiens		
US-08-913-832A-2		
Query Match	8.3%: Score 241; DB 4; Length 1912;	
Best Local Similarity	23.1%: Pred. No. 4e-11;	
Matches 97; Conservative 51; Mismatches 172; Indels 100; Gaps 14;		
QY 129 PTKRKASEEAKAAAPALTPRGTAAPGSGQLAKPKPESSAEOQLPLGNGIQMTASAV 188		
DB 217 PFKSGSASVAAAAAIVAVESWVATEV-APPPVEVPIRKAKTEGKGPNA----- 270		
QY 189 QRAVAMSSGVPKARGAVEGLILIOVFESG-----SKCIVGGFFTPSPFED----- 238		
DB 271 -RRPKSPRPVPAKKKPKVAPLTKLIGFGSKRKSSSEDDLDVESDFDMSINSY 329		
QY 239 SGSGKNKARSSSGPKPVYRAKGAAGGAEARLGOOGSVAPALPSPDLHOKNDE 298		
DB 330 SVDSGSTRSRSSRKRLRTK-----KKKGEE-EYAVDGYETD-----HODY 372		
QY 299 CAVCRDGGELICDGPRAFLACISPLREIPSTGWRSSCLQTVQVOPRAEPPRQ 358		
DB 373 CEVCOQGEIILCTPCRAYHMVCLDPMERAKPEKWKSPCEK---BGIOWEAKED--- 426		
QY 359 EPPVEFTPLPGLRAGEVRGPPGEPLAGMDTTLVYKHLPPPSAAPLPGIDSSSLHL 418		
DB 427 -----NSGEILIEVVGDLLEEDD-----HHMER----- 451		
QY 419 CVPREGQNLAPGARCVCVGDGTDVLRCTHCAAFHMRCHFPAGTSRGTGLRCRSCGD 478		
DB 452 -----CVCCKDGGELLCDDTCPSVHHCCLMPLEIRNGEMLCRCRCIP 496		
QY 479 VTPAPEGVLT-----APSPARL-APGPAKDDTASHEPALHRDLESLSLSEHTFDGI 531		

Db 497 ALKGKVKILIMKQGPSPPTVPRPPDADPNTSPKPLEGRERORFVK---WQGNSTW 553

RESULT 2

US-09-249-181A-2
Sequence 2, Application US/09249181A
Patent No. 6440679
GENERAL INFORMATION:
APPLICANT: Seelig, Hans Peter
APPLICANT: Renz, Manfred
TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
FILE REFERENCE: 8484-0059-999
CURRENT APPLICATION NUMBER: US/09/249,181A
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 08/913,832
PRIOR FILING DATE: 1998-01-12
PRIOR APPLICATION NUMBER: PCT/DE96/00444
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 1912
TYPE: PRT
ORGANISM: Homo sapiens
US-09-249-181A-2

Query Match

Best Local Similarity 8.3%; Score 241; DB 4; Length 1912;
Matches 97; Conservative 51; Mismatches 172; Indels 100; Gaps 14;

QY 129 PTRKASEEAAAPALTPRGTAQSGQLAKPKPKPESSAEOQLPLGNGIQTMSASV 188
Db 217 PFKSSGSAVAAAAYAVESMTATEV-APPPEVEYIRKKTKEGKPN- 270
QY 189 QRAVAMSSGDVPGARGAVEGLIDQVESSG---SKKCIQVGEFFYPSKFE- 238
Db 271 -RRKPKGSPVPPDKKPKPKKAPLKLKIGFGSKRRKSSSEDDLDVESPFDASINSY 329
QY 239 SSGSKNKARSSSGPKPLVRAKGAAGAPGGEARLGOOGSVAPALPLPSDQHLKND 298
Db 330 SVSDGSTSRSSRRKRTTK-----KKKKEE-EYTAVDGYETD-----HODY 372
QY 299 CAVCRGGELICDCCPRAHFLACSLPRLREIPSGTWRCSSCLQATVOEVPRAEPPRQ 358
Db 373 CEVQGGEGEILLCDTPRAYHMYCLDPMKAPRGKWCSPHCEK---EGIQWEAKED--- 426
QY 359 EPPVEPLPLGRSAGEEYVGPGEPLAGMDTLVYKHLPAFSAAPLPLGLDSSALHPL 418
Db 427 -----NSEGEELIEVGGDLEEDD---HHMEF----- 451
QY 419 CVGPESQMLAPGARGVCGDGTDLRCHCAAFHMRCHFPAGTSNRGTGLRCSCSD 478
Db 452 -----CRVCKDGGELLCDCTCPSSYHHCINPLPLPELPCNEMWLCPRCTCP 496
QY 479 VTPAVVEGL-----ASPARL-APGPAKDDTASHEPALHRDLSLSEHTFDGILW 531
Db 497 ALKGKVKILIMKQGPSPPTVPRPPDADPNTSPKPLEGRERORFVK---WQGNSTW 553

RESULT 3

PCT-US91-06418-1
Sequence 1, Application PC/TUS9106418
GENERAL INFORMATION:
APPLICANT: Oklahoma Medical Research, Foundation, et al
TITLE OF INVENTION: Antigens Associated with Polymyositis
TITLE OF INVENTION: and with Dermatomyositis
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Kilpatrick & Cody
STREET: 100 Peachtree Street
CITY: Atlanta
STATE: Georgia

COUNTRY: US
ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US91/06418

FILING DATE: 19910905

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/579023

FILING DATE: 09-JUL-1990

ATTORNEY/AGENT INFORMATION:

NAME: Padst, Patrea L.

REGISTRATION NUMBER: 31,284

REFERENCE/DOCKET NUMBER: OMRI120

TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-572-6508

TELEFAX: 404-572-6555

SEQUENCE CHARACTERISTICS:

LENGTH: 351 amino acids

TYPE: AMINO ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

ORIGINAL SOURCE:

ORGANISM: Homo sapien

TISSUE TYPE: Sera

IMMEDIATE SOURCE:

CLONE: L1

PCT-US91-06418-1

Query Match 8.1%; Score 236.5; DB 5; Length 351;
Best Local Similarity 25.6%; Pred. No. 1e-11;
Matches 84; Conservative 28; Mismatches 131; Indels 85; Gaps 11;

QY 218 GSKKCIQVGEFFYPSKFE---DSGSKNKARSSSGPKPLVRAKGAAGAPGGEAR-L 273
Db 8 GGRKRGSSDDEGPEEABESDLSGVSASGRPDGP---VTKTKLRERPKKKKVL 64
QY 274 GQOGSVAPALPLPSDQHLKNDKCAVCRGGELICDCCPRAHFLACSLPRLREIPSG 333
Db 65 G-----CPVAGGEYVDGYETDHDQYCEVCGGGEITLCTCPRAYHLYCLDPELDRAPEG 120
QY 334 TWRCSSCLQATVOEVPRAEPPRQEPPEVETPLPLGRSAGEEVRGPGEPLAGMDTLV 393
Db 121 KWSCHKEKEGYWAKKEEYEE-----GEE-BEKEEE---DHME 161
QY 394 YKHLPAFSAAPLPLGLDSSALHPLLCVGPESQMLAPGARGVCGDGTDLRCHCAAF 453
Db 162 Y-----CRVCKDGGELLCDCTCPSSYHHCINPLPLPELPCNEMWLCPRCTCP 496
QY 454 HMRCHFPAGTSRPGTGLRCSCSDGYTPAVVEGL-----ASPARLAPGPAKD 503
Db 183 HIHCLNPLPLPIPGNEMWLCPRCTCPVLRGVOKILHMRGEPVAVPAQADGNP---D 239
QY 504 TASHEPALHRDLSLSEHTFDGILW 531
Db 240 VPPRPLGRSEREREFVK---WVGLSTW 264

RESULT 4

US-08-942-686-2
Sequence 2, Application US/08942686
Patent No. 6183988
GENERAL INFORMATION:
APPLICANT: Bloch, Donald B.

Db 335 -----SAP-----APKGRSPSPGSS----- 351

QY 451 AAFHMRCHPAGTS-RP-GTGLRCRSCSGDVTAPAVEGVLAAPSPARLAPGPADDTASHE 508
 Db 352 -----PSGRVAKVESPEGVARRR-----VSPVPOSGRI--TPPRRASP--DGFSPPS 395

QY 509 P 509

Db 396 P 396

RESULT 6
 US-09-579-181-2
 ; Sequence 2, Application US/09579181
 ; Patent No. 6365372
 ; GENERAL INFORMATION:
 ; APPLICANT: Chiviva, John
 ; APPLICANT: Yaciuk, Peter
 ; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
 ; FILE REFERENCE: 16153-4247
 ; CURRENT APPLICATION NUMBER: US/09/579,181
 ; PRIOR FILING DATE: 2000-05-25
 ; PRIOR APPLICATION NUMBER: 60/136,620
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 2
 ; LENGTH: 2972
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-579-181-2

Query Match 5.3%; Score 154; DB 4; Length 2972;
 Best Local Similarity 22.0%; Pred. No. 0.001;
 Matches 114; Conservative 44; Mismatches 195; Indels 166; Gaps 24;

QY 102 KDVDLSQPRKGRKPPAV-----PK-ALVP-PPRLPTKRKASEEARAAPAA 145
 Db 1027 ROLAVGQPRLOMPPTVNNNTGVYKIVROAPRDGLTPVPLAPAPRPPS-----SGLPAV 1082

QY 146 LTPRGTAPEGSQLKAKPKPKPESSAEQRLPLNGIOTMSASVORAVAMSSGDV-----PG 201
 Db 1083 LNRPLTLTPG-----RLPPTLTGTARAPMPT--PTLVRLPLKLIVHSPPEVSASAPG 1132

QY 202 ARGAVEGILIQVEFESGSKKCIQVGEFTYTPSKFEDSGSGKKNARSSGPKPLVRAKGA 261
 Db 1133 AAP-----LTSSPLHVPSSLPGPASPMPIMPSS--PLASPVSS 1170

QY 262 QGAAPGGGARLGGQGSVPAPLALP-----SDQLHQKNEDECAVCRDGGELICDGCPR 316
 Db 1171 TVSVPLSSSLPISVPTTLAPASAPLTIPIISAP-----LTVSASGAL----- 1213

QY 317 AFHLACLSPPLREI-----PSGTWRCSSCIQATVOEVOAPRAEPPPOEPVETP 365
 Db 1214 ---LTSVTPPLAPVYPAAGPPSLOPSGASPSASALTGLGATAPSLSSSOTPGHPLLAP 1270

QY 366 LP-----PGLRSAGEEVRGPGEPLAGMDTTLVYKHLAP-----P 401
 Db 1271 TSSHVPGLNSTVAPACSPVLVPAASALASPPAPNPAPAPQASILAPASSASQALATPLAP 1330

QY 402 SAAP-----LPGIDSSALHPLLCVGPPEGQONLAPGARCGVCGDGTDLRCHCAAAAFHWRK 457
 Db 1331 MAAPQATILAPSPAPPLAPLPLVLA-----SPGAAPVLASSQTPV----- 1370

QY 458 HFPAGTSRPGTGLRCRSCSGDVTAPAVEGVLAAS-----PARL-----ADGPAKDDTASHE 508
 Db 1371 PVMAPSSTPGTSL-----ASASVPAAPTP-VLAPSSQTQTLMLPAPVSPPLSPASTQTALAA 1425

QY 509 PALHRDLESLSLSEHTFDGILQWAIQSM--ARPAAPPS 545
 Db 1426 PAL-----APTLCGSSSPSQTLSLGTGNGPGPFPT 1454

RESULT 7
 US-09-579-181-1
 ; Sequence 1, Application US/09579181
 ; Patent No. 6365372
 ; GENERAL INFORMATION:
 ; APPLICANT: Chiviva, John
 ; APPLICANT: Yaciuk, Peter
 ; TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
 ; FILE REFERENCE: 16153-4247
 ; CURRENT APPLICATION NUMBER: US/09/579,181
 ; PRIOR FILING DATE: 2000-05-25
 ; PRIOR APPLICATION NUMBER: 60/136,620
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 3118
 ; TYPE: PRT
 ; ORGANISM: Human
 US-09-579-181-1

Query Match 5.3%; Score 154; DB 4; Length 3118;
 Best Local Similarity 22.0%; Pred. No. 0.0011;
 Matches 114; Conservative 44; Mismatches 195; Indels 166; Gaps 24;

QY 102 KDVDLSQPRKGRKPPAV-----PK-ALVP-PPRLPTKRKASEEARAAPAA 145
 Db 1173 ROLAVGQPRLOMPPTVNNNTGVYKIVROAPRDGLTPVPLAPAPRPPS-----SGLPAV 1228

QY 146 LTPRGTAPEGSQLKAKPKPKPESSAEQRLPLNGIOTMSASVORAVAMSSGDV-----PG 201
 Db 1229 LNRPLTLTPG-----RLPPTLTGTARAPMPT--PTLVRLPLKLIVHSPPEVSASAPG 1278

QY 202 ARGAVEGILIQVEFESGSKKCIQVGEFTYTPSKFEDSGSGKKNARSSGPKPLVRAKGA 261
 Db 1279 AAP-----LTSSPLHVPSSLPGPASPMPIMPSS--PLASPVSS 1316

QY 262 QGAAPGGGARLGGQGSVPAPLALP-----SDQLHQKNEDECAVCRDGGELICDGCPR 316
 Db 1317 TVSVPLSSSLPISVPTTLAPASAPLTIPIISAP-----LTVSASGAL----- 1359

QY 317 AFHLACLSPPLREI-----PSGTWRCSSCIQATVOEVOAPRAEPPPOEPVETP 365
 Db 1360 ---LTSVTPPLAPVYPAAGPPSLOPSGASPSASALTGLGATAPSLSSSOTPGHPLLAP 1416

QY 366 LP-----PGLRSAGEEVRGPGEPLAGMDTTLVYKHLAP-----P 401
 Db 1417 TSSHVPGLNSTVAPACSPVLVPAASALASPPAPNPAPAPQASILAPASSASQALATPLAP 1476

QY 402 SAAP-----LPGIDSSALHPLLCVGPPEGQONLAPGARCGVCGDGTDLRCHCAAAAFHWRK 457
 Db 1477 MAAPQATILAPSPAPPLAPLPLVLA-----SPGAAPVLASSQTPV----- 1516

QY 458 HFPAGTSRPGTGLRCRSCSGDVTAPAVEGVLAAS-----PARL-----ADGPAKDDTASHE 508
 Db 1517 PVMAPSSTPGTSL-----ASASVPAAPTP-VLAPSSQTQTLMLPAPVSPPLSPASTQTALAA 1571

QY 509 PALHRDLESLSLSEHTFDGILQWAIQSM--ARPAAPPS 545
 Db 1572 PAL-----APTLCGSSSPSQTLSLGTGNGPGPFPT 1600

RESULT 8
 US-09-051-019-2
 ; Sequence 2, Application US/09051019
 ; Patent No. 6103229
 ; GENERAL INFORMATION:
 ; APPLICANT: KAHMANN, Regine and QUADBECK-SEEGER, Claudia
 ; TITLE OF INVENTION: Regulatory gene from *Ustilago maydis*
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Kell & Weinkauff
 ; STREET: 1101 Connecticut Avenue

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage
COMPUTER: IBM AT-compatible, Pentium processor
OPERATING SYSTEM: Windows 98
SOFTWARE: Wordperfect version 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/051,019
FILING DATE: 31-MAR-1998
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2289 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-051-019-2

Query Match 5.2%; Score 151.5; DB 3; Length 2289;
Best Local Similarity 22.1%; Pred. No. 0.0012;
Matches 84; Conservative 41; Mismatches 118; Indels 137; Gaps 19;

QY 28 LHALADHDVPEDEKQETLHLKEKGCQPAFHALLSMLTQDSTALIDFWRVLFKDYNL 87
DB 285 LINSLS-ADARASQNYOEOL-----OKFHA----- 308
QY 88 ERYGRLO--PILDSFPKVDLSQPR-----KGRKPPAVPKAL---VPPPR 127
DB 309 -QGGKKRVSVPYID--RSVDLYQLKLVISLGGYDAVCAKRRKMSDTRKIGYSDKEGQ 365
QY 128 LPTKKRAS-----EARRAAPALTPRGTAAPSGSQLAKRPPKPPSSAEQRLPLGN 179
DB 366 LSTQYKAAVTRIIIPPEEFLLAKKEQSRPNG-SSVSPQLAQSALMGATASTDQE---N 420
QY 180 GIO--TMSAYGAVAMSSGDVPGARGAVGILIQVFEESGSKKCIQVGEFTIPSKFE 237
DB 421 GVKHPSMSPLD--AAPSGD-----AGHFHFKPEPFT 450
QY 238 DSGSGKNKRRSS-----SGPKPLVRAKGAOGAAPGGGEARLGGQGSVP-APL 283
DB 451 AAGAALANATPVLLETPTQSPSTVASTRSARAKRSATSTPASSSRSLDITSTPMPL 510
QY 284 A-----LPSDPOL-----HQRNDE-CAVC--RDGELICDGCPRAPHILACLS 324
DB 511 ISRRKGVSPHLEADSYLLAQAGNQAEQMCETCLRGEDGPNMLLDCENRGYHMYCLQ 570
QY 325 PPLREIPSGTWRCCSCLQAT 344
DB 571 PALTSPKSWQFCPCLVGT 590

RESULT 9
US-08-976-255-11
Sequence 11, Application US/08976255
Patent No. 6136581
GENERAL INFORMATION:
APPLICANT: Jono, Keith E.
APPLICANT: Plozman, Gregory
TITLE OF INVENTION: KINASE GENES AND USES
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage

COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSeq for Windows 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,255
FILING DATE: No. 6136581ember 21, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/031,675
FILING DATE: No. 6136581ember 22, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 229/182
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1384 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-976-255-11

Query Match 5.1%; Score 149; DB 4; Length 1384;
Best Local Similarity 23.6%; Pred. No. 0.00097;
Matches 96; Conservative 42; Mismatches 141; Indels 128; Gaps 21;

QY 101 PKVDLSQPRKRPKPPAVPKALVPPRLPTRKASEEARA-AAPALTFRGTASPSQQL 159
DB 790 PK--LATEAGTGPRLPLRPSVSPSGEAPLSEEAAPAPADALPSPPTPATGEVS 846
QY 160 A-----KPKPPSSAEQRLPLGNGIOTMSA-----SVQRA 191
DB 847 AIKLASALNGSSSPSEVEAPSSSEDEDTAEAVSGIPTDSSDGLQARRPDVPAFRSLQK 906
QY 192 V---AMSSGDVPGAR--GAVGILIQVFEESGSKKCIQV--GEFTIPSKF--EDSGS 241
DB 907 VGTPTDLSIDTIDPSASDGEYFSPSATGSGOPRALDSYDTEVESPEVLLKEAD 966
QY 242 G-----KNKARSSSGKPLVRAK-----GAOGAAGGGEARLG 274
DB 967 GCEPQAFELASGEGERPRLSTSLGMEKNRYRSATVSDLEAEALASGPKKCG 1026
QY 275 QQGSVAPALPLPSDPOLHQRNDECAVCRDG-----GELICDGCPRAPHILACLS 324
DB 1027 GDRAPGPELGLPS--TGQPSQYVC--LRPGVSGEAGSGPGEVL----- 1066
QY 325 PPLREIPSGTWRCCSCLQATVQEQPRAEP--RQPEPP-----VETPLPGILRSAG- 374
DB 1067 PPLQLLEGSSPEPSTCPSGLVPE--PPEQGPAAKVPSPSCSOFLLTPVP--LRSEGN 1123
QY 375 -EEVRGPPG-----EPLAGMDTTLVYKHLPAAPPASAPLPGLDSS 412
DB 1124 SSEFGQPGLLSGPAPQKRMGGPCT-----PRAPLALPLGLPRAA 1163

RESULT 10
US-08-545-860D-55
Sequence 55, Application US/08545860D
Patent No. 6040140
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaan, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Regio
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &

```

Query Match           5.1%; Score 147; DB 3; Length 1093;
Best Local Similarity 18.7%; Pred. No. 0.001;
Matches 113; Conservative 77; Mismatches 249; Indels 166; Gaps 20;

QY      38 VPEKRFQETLHLKE-----KEGRPOAFH---ALLSWLLTODSTALIDF 77
          |||::||::|       :||| | | |   |::||::|::|
Db      111 VPHDRENTTCYCIECTEGRESKAASGACMTCNHNCRGFHFVYCAOMAGILCEE--VLEV 168
          |||::||::|       :||| | | |   |::||::|::|

QY      78 WRVLFKDYNLERYGLDPLDS-----PRKVDL 106
          |||::||::|       |

Db      169 DNIVVCYGCKAHFYFSMKTSRRSHSGGGGGGAGGGGSMGGGSGCFTSGRRSRASSTQDE 228
          |||::||::|       |

QY      107 SQP-----RKGRKPRAVKALVPRLPTKKRASEEARAARAAVALTP 148
          |||::||::|       |

Db      229 KHPTHNEGOKKSRRDKERLKOKHHKRREPSPSITLPVVVLADRVKSSASSSSHHSEAST 288
          |||::||::|       |

```

```

OY      149  RGAASGSOAKAPPPKPFESSAEQOURLPLNGIDTMSASVQRAYAMSGDVPGARGAVEG  208
Db      289  QETSESREKGGKSSSHSLSHKCKKLSSCGVSTFSASSSSSSSSSGGPPQPAVSS  348
OY      209  I-----LIQVFESGGSKKCIQVGGEFYTPS-----K  235
Db      349  LOSPPDFSAFPLEQPEEDYYKPTAPAPASAPSPSAPEPPKADLEQKVVSCHGPIMR  408
OY      236  FEDSGSGKKMKARSSSG---FKPLVRAGAGAAAGGGEARLGQOGSVAPRLALPSDPLH  292
Db      409  FSTTTTSSCGRARPSPGDKYSPVHTSGAS---AGTHKKRPALSAATPVPADETPETGLK  464
OY      293  QKNEDECAVCROGELICDCGCPRAFH-----LACTLPLREIPSGT-----WRCSS  339
Db      465  EKKHKASKRSRHG-----PGRPKGSNKKECTGGPAAPSLPSQDLAFTTALASPFSGGS  518
OY      340  CLOATYQVQAPRAEERPOEP--PVETP-LPPLGRSAGEEYRGPPGPEPLACMDTTLVYKH  396
Db      519  LVSSGGGGLSRFFPGSGSLPSLSTLSPLLGAGIYTSNKPDISHSQGLRAVCSTPLSSS  578
OY      397  LPAPSAAPLPLGDSLALHPLLCVGEQGQONLAPGACGCGSDGTDLRCHHCAAFHWR  456
Db      579  LLGPQTSALPLRSR---PFTSTLPSSASISITTOYFSLAG-STESLPSTHIFGT----  630
OY      457  CHEPAQTSRP-----GTGLRCRSCSGDVTPAPVEGYLAPSVARLAPSPAK  501
Db      631  ---PMGAVNPLLSQAESSHTPELDECSFCRCRTS-----PQSLSMSPISSLPA-LF  680
OY      502  DDTAS 506
Db      681  DOTAS 685

RESULT 11
PCT-US94-004496-55
: Sequence 55, Application PC/TUS94004496
GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaanl, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
TITLE OF INVENTION: For Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the ALL-1
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
ADDRESSEE: Norris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/004496
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deluca Esq., Mark
REGISTRATION NUMBER: 33,229
REFERENCE/DOCKET NUMBER: TJU-1242
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 1093 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

```


COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/818,823
 FILING DATE: 14-MAR-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/396,479
 FILING DATE: 02-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Osman, Richard A
 REGISTRATION NUMBER: 36,627
 REFERENCE/DOCKET NUMBER: A-59450-1/RAO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 494-8700
 TELEFAX: (415) 494-8771
 TELEX: 210 277299

INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 902 amino acids
 TYPE: amino acid
 TOPOLOGY: linear

MOLECULE TYPE: protein
 US-08-818-823-6

Query Match 5.0%; Score 144.5; DB 1; Length 902;
 Best Local Similarity 23.7%; Pred. No. 0.0013;
 Matches 113; Conservative 29; Mismatches 136; Indels 199; Gaps 25;

```

QY 113 RKPAPVAKLVPPRLPTKRKASEARAAAPALTPRCTASPGSQ-----LKA 160
DB 63 RPPPRGMHSPPR-PAPSPGTESQPARSVRLGGGAGAGAGGVLECPSTRITSI 121
QY 161 KPKKPPSSAEQRLPLGNGIQTMSASVQRAVAMSSGDVPGARGAEGILIQVFEESGS 220
DB 122 SPPEPPALEDNPDWDG-----SPRDYPPPEG-----FCGY 155
QY 221 KKC-IQVGEFTYPSK-----FEDSG-----SGKKAKSSSG---PK 253
DB 156 REAGAGGGGAFSPSSGSSLSKMSFTSDASDEALTYACDEVESELENAASRGGLGSP 215
QY 254 PLVRAK-----GAQGAAPG-----GEAR-----272
DB 216 PSPRASPRWTPEDPWSTLVGSPGGRPEDESWLLSAPGTPPASPRAAPGCKRYSSG 275
QY 273 -----LGGQGS-----VPAPLALPSPDLHOKNDECAVCARDGGLICDGC 315
DB 276 TSSASPAISRSGISLEESSEPPPPPLPLADP-----GSP 312
QY 316 RAFLHACLSPPLREIPSGTWRCSSCLOATVQEVQRAPEERPOEPVETPLPGLBSAGE 375
DB 313 GFEDVYG-APPAESTIQKRRITSS-EQAV---ALPREEP---ASNGKLPGLG---AE 359
QY 376 EYRGPPG---EPLAGMDTLVYKHLPPAP--SAAPLPG---LDSALHPLLCGPEGOQ 426
DB 360 ESVAPPGSGSRKEVAGMD---YLAVSPPLAMSKARIGHSPIRTGALLPLDMLPISQYE 415
QY 427 NLAPGARGCGVDGTDVLCHTCAAFHMRCHFPACTSRPGTGLRCRSCGSDVTPAP 483
DB 416 QLE-----LRIEVOPRANH-RAHYETEGSR-----GAVKAP 446

```

RESULT 14
 US-09-370-838-216
 ; Sequence 216, Application US/09370838
 ; Patent No. 6444425
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Mohamatch, Roadoh

APPLICANT: Secrist, Heather
 TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
 TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
 FILE REFERENCE: 210121.475C1
 CURRENT APPLICATION NUMBER: US/09/370,838
 EARLIER FILING DATE: 1999-08-09
 EARLIER APPLICATION NUMBER: US 09/285,323
 NUMBER OF SEQ ID NOS: 289
 NUMBER OF SEQ ID NOS: 289
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 216
 LENGTH: 527
 TYPE: PRT
 ORGANISM: Homo sapien
 US-09-370-838-216

Query Match 4.9%; Score 143; DB 4; Length 527;
 Best Local Similarity 23.8%; Pred. No. 0.00085;
 Matches 126; Conservative 36; Mismatches 191; Indels 176; Gaps 29;

```

QY 105 DLQPRKG-----RKPAPVAKLVPPRLPTKRKASEARAAAPALTPRCTASPG 155
DB 30 ELTPPREKALENGELRSPDAGEKVLVNGILTPK---SEDKVSENGLAEPRTTERP 86
QY 156 SOLKAKP-----KPPSSAEQRLPLGNGIQTMSASVQRAVAMSS-----CD-VPGAR 203
DB 87 ETGPMWRAPGMEKTPESWGAPPT---GEPAPETSLERAPAPSAVSSRNGGETAGP 142
QY 204 GAV-EGILIQVFEESGSKKIQVGEFTYPSKF-DSSGKKAKKASSGPKLVRAKG 260
DB 143 GRAPKNGTL-----EGETERAPETGAPRAPAGAGRIDLSG-----RAPV 184
QY 261 AGGAAPGG-----EARLGGQGSV---PAPLALPSD-----POLHOKNE 296
DB 185 GTGTAPGGGPGSGVDKAGVNDTRPQPPPLPPEEAPRLRPPAPRAPEVAPEGE 244
QY 297 DECAVORDGE-LICDGE-----CPRAFLACLSPPLREIPSGTWRCSSCLOAT 344
DB 245 PGAPDSRAGGDVALSGDGPPEPKERKGPMPRLF-LDLGP-----QGN 287
QY 345 VOEVQRAPEERPOEPVE-TPLP-PGLR-----SAGEVNGPPEPLAGMDTL 392
DB 288 SQQIKARLSRLALPPLTLTPPGGPRRPPEWEGADAGAGGAG-----AG-----336
QY 393 YKHLPPAPSAAPPLGLDSSALHPLLCVPGEQNLAPGARGVCGDGT-----VLR 445
DB 337 -----APPAEEDGEDDEDE-----EEDDEAAPGAAGPRGPRARAAPVYVS 383
QY 446 CTHCAAFHMR---CHFPAGTSRPGTG-----LRCNSCGDVT-----PAP 483
DB 384 SADADAAPRLKRLKSPKADPEDESELERKKRMVSFHGDVYVLPDQETPTNELSVQAP 443
QY 484 VEGVLAPSPARLAPPAKDDTASHHEPALHRLDLESLSHTDGLIOMA 532
DB 444 PEGDIDPESTPAPPTPP-----HPATPGDGPPS--NDSGFGSGFEWA 483

```

RESULT 15
 US-08-246-489-2
 ; Sequence 2, Application US/08246489
 ; Patent No. 6225049
 ; GENERAL INFORMATION:
 ; APPLICANT: Lan, Michael S.
 ; APPLICANT: No. 6225049kins, Abner L.
 ; TITLE OF INVENTION: NOVEL HUMAN INSULINOMA-ASSOCIATED CDNA
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear
 ; STREET: 620 Newport Center Drive
 ; CITY: Newport Beach
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 92660

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/246,489
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/901,715
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Israel, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH012,012A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 510 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-246-489-2

Query Match 4.9%; Score 142.5; DB 4; Length 510;
Best Local Similarity 22.8%; Pred. No. 0.00089;
Matches 118; Conservative 41; Mismatches 175; Indels 183; Gaps 31;

OY 114 KPPA---VKALVPPRLTKKASEEARAAAPALT-----PRG---TASPSQLKA 160
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 42 EPPAPSPVGPPLPP-----PAERAHALAAALACAPGPPPGPPRAHFGNPEAA 94
OY 161 KP-----PKPSSAEQOR-----LPLGNGIQTMSASVORAVAMSSGDVPGAGAVEGI 209
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 95 HPAPLYSPTRPYREHEKHYPERSFNLSGVSASFPTPAL-LGGGGGGASGA---- 149
OY 210 LIQOVESGSKKCIQVGE--FYTPSKFEDSGSKKAKRSSGKRP-----LYRAK 259
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 150 -----GGGGTC---GGDPLFAPAEELKMTAFSAGAEANGPGPPLPPAALLRPP 198
OY 260 GAOGAAPGGE--ARLGQGSVPAPLAPSDPOLHOKNEDEC----- 299
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 199 GKRPPTTAEPKAPKAVKAPGAKPKAT--RKLH--FEDEVTTSPVLGLKIKGVEAP 253
OY 300 -----AVCRDGEELICDCGPRAF--HLACLSPPLREIPSGTWRCSSCLQ-----ATVOE 347
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 254 RCRAGGAAPRLGEFI-CQLCKEYADPPALAQHKCSRIVVEYRCPKCAKVFSCPANLAS 312
OY 348 VOPRAEPPRPQEPVTPLP-PGLRSAGEVRG-----PPGEPLAGMDTTL----- 392
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 313 -HRRMKRPPRAAPAAARAPPEAAARAAREAPGSGSDRDTSPSGVSESGEDGLYCHH 371
OY 393 -----YKHLPAAPSA-----APL--PGIDSSALHPLLCVPE----- 423
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 372 CAKKRRQAYLKRHLAHQALQAKGAPLAPPAEDLALYF-----GPDEKAPQEAAGDE 427
OY 424 --GOONLAPGAR--CGVCGD-----GTDVLRCTHCAAAFHWRCHPAG 462
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 428 GAGVILGLSASAECHLCPCVSGESFASKAQDERHLRLHAQVFPCKYCPATFY----- 479
OY 463 TSPRTGLRCRSGSDVTPAPVGE--VLAPSPARLA 496
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB 480 -SSPGLTRHINKCH-----PSENROYILLQVVRPA 509
```


GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 13, 2003, 17:50:47 ; Search time 14 Seconds

(without alignments)
1614.614 Million cell updates/sec

Title: US-09-509-595b-2

Perfect score: 2902
Sequence: 1 MADDAALRLRLRLHRTETAV.....DGLIOWAIOSMARPAAPFPS 545

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2902	100.0	545	1 AIRE_HUMAN	043918 homo sapien
2	2005	69.1	552	1 AIRE_MOUSE	092063 mus musculu
3	270.5	9.3	1944	1 CHD3_HUMAN	012873 homo sapien
4	241	8.3	1912	1 CHD4_HUMAN	014839 homo sapien
5	228	7.9	689	1 SP11_HUMAN	09hb58 homo sapien
6	227	7.8	1050	1 SP11_HUMAN	015164 homo sapien
7	227	7.8	1982	1 CHDM_DROME	097159 drosophila
8	217	7.5	1051	1 TP1A_MOUSE	064127 mus musculu
9	206	7.1	1787	1 CHD3_MOUSE	022516 caenorhabdi
10	204.5	7.0	882	1 LY10_HUMAN	013342 homo sapien
11	204	7.0	1127	1 TP1G_HUMAN	09upn9 homo sapien
12	186.5	6.4	607	1 YAZ7_SCHPO	009698 schizosacch
13	179.5	6.2	404	1 YAC5_SCHPO	009819 schizosacch
14	168.5	5.8	660	1 YH11_EBV	P03181 Epstein-Bar
15	161	5.5	684	1 YH11_EBV	004779 saccharomyc
16	160	5.5	835	1 TP1B_HUMAN	013263 homo sapien
17	159.5	5.5	1083	1 TP1B_HUMAN	000268 homo sapien
18	157	5.4	834	1 TP1B_MOUSE	062318 mus musculu
19	152.5	5.3	2161	1 SHK1_HUMAN	09y566 homo sapien
20	152	5.2	366	1 REO2_XENLA	09y566 xenopus lae
21	152	5.2	380	1 REO2_XENLA	P58267 gallus gall
22	152	5.2	810	1 FAL2_HUMAN	012830 homo sapien
23	152	5.2	1049	1 CAL3_BOVIN	P04258 bos taurus
24	151.5	5.2	728	1 YJ89_YEAST	P47156 saccharomyc
25	149.5	5.2	2167	1 SHK1_RAT	015654 homo sapien
26	147.5	5.1	476	1 TRX2_HUMAN	09umne homo sapien
27	147.5	5.1	2715	1 AP17_HUMAN	P55198 homo sapien
28	146	5.0	1093	1 NCR2_HUMAN	09y618 h nuclear r
29	145	5.0	2517	1 NECA_HUMAN	014934 homo sapien
30	144.5	5.0	902	1 DRPL_RAT	P54258 rattus norv
31	144.5	5.0	1183	1 REO1_XENLA	09y638 xenopus lae
32	144	5.0	388	1 PRGR_MOUSE	000175 mus musculu
33	143.5	4.9	923	1	

34	143.5	4.9	1560	1 SMCX_HUMAN	P41229 homo sapien
35	143	4.9	427	1 REOC_CHICK	P58270 gallus gall
36	143	4.9	1267	1 Y211_HUMAN	092610 homo sapien
37	143	4.9	1554	1 SMCX_MOUSE	P41230 mus musculu
38	142.5	4.9	510	1 TAL_HUMAN	001101 homo sapien
39	142	4.9	775	1 ICP0_HSV11	09y666 herpes simp
40	141.5	4.9	1539	1 SMCY_HUMAN	09y666 homo sapien
41	141	4.9	2142	1 BAT2_HUMAN	P48634 homo sapien
42	140.5	4.8	1548	1 SMCY_MOUSE	062240 mus musculu
43	140	4.8	1411	1 TCOR_HUMAN	013428 homo sapien
44	140	4.8	3866	1 HRX_MOUSE	P55200 mus musculu
45	139.5	4.8	1214	1 BRFL_HUMAN	P55201 homo sapien

ALIGNMENTS

RESULT 1
AIRE_HUMAN STANDARD: PRT; 545 AA.
AC 043918: 043932: 043922: 075745;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Autoimmune regulator (Autoimmune polyendocrinopathy candidiasis
DE ectodermal dysplasia protein) (APECED protein).
GN AIRE OR APECED.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Thymus;
RC MEDLINE=98061086; PubMed=9398839;
RX Nagamine K., Peterson P., Scott H.S., Minoshima S.,
RA Heino M., Krohn K.J.E., Laiotti M.D., Mullis P.E., Antonarakis S.E.,
RT Kawasaki K., Asakawa S., Ito F., Shimizu N.,
RL "Positional cloning of the APECED gene.";
Nat. Genet. 17:393-398(1997).
[2]
SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Thymus;
RX MEDLINE=98061087; PubMed=9398840;
RA Aaltonen J., Björnsen P., Perheentupa J., Horelli-Kuitunen N.,
RA Palotie A., Peltonen L., Lee Y.S., Francis F., Hennig S., Thiel C.,
RA Lehnach H., Yaspo M.-L.;
RT "An autoimmune disease, APECED, caused by mutations in a novel gene
featuring two PHD-type zinc-finger domains.";
Nat. Genet. 17:399-403(1997).
[3]
SEQUENCE FROM N.A.
RA Lee Y.S., Francis F., Hennig S., Thiel C., Reinhard R., Lehnach H.,
RA Yaspo M.-L.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Pak H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordiek G., Hornischer K., Brandt P.,
RA Scharte M., Schoen O., Desario A., Reichelt J., Kauer G., Blocker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesseltmann L., Dagand E.,
RA Wehrmeyer S., Borzjym K., Gardiner K., Nizetic D., Francis F.,
RA Lehnach H., Reinhardt K., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21.";
Nature 405:311-319(2000).
[5]

RP SUBCELLULAR LOCATION.
 RX MEDLINE-99135905; PubMed-9931333;
 RA Björres P., Pelto-Huikko M., Kaukonen J., Aaltonen J., Pelttonen L.,
 RA Umanen I.;
 RT "Localization of the APECED protein in distinct nuclear structures.";
 RL Hum. Mol. Genet. 8:259-266(1999).
 RN [6]
 RP PARTIAL SEQUENCE, SUBUNIT STRUCTURE, DNA-BINDING, AND PHOSPHORYLATION.
 RX Kumar P.G., Laloraya M., Wang C.-Y., Ruan Q.-G., Davoodi-Semlromi A.,
 RA Kao K.-J., She J.-X.;
 RT "The autoimmune regulator (AIRE) is a DNA-binding protein.";
 RL J. Biol. Chem. 276:41357-41364(2001).
 RN [7]
 RP SUBCELLULAR LOCATION, TRANSCRIPTION ACTIVATION, AND MUTAGENESIS OF
 RX LEU-28: CYS-302 AND CYS-437.
 RA Pitkanen J., Vaehemurto P., Krohn K.J.E., Peterson P.;
 RT "Subcellular localization of the autoimmune regulator protein.
 RT characterization of nuclear targeting and transcriptional activation
 RT domain.";
 RL J. Biol. Chem. 276:19597-19602(2001).
 RN [8]
 RP VARIANT APECED PRO-28.
 RX MEDLINE-99103466; PubMed-9888391;
 RA Heino M., Scott H.S., Chen Q., Peterson P., Maenpää U.,
 RA Papasavvas M.-P., Miltaz L., Barras C., Rossier C., Chrousos G.P.,
 RA Stratakis C.A., Nagamine K., Kudoh J., Shimizu N., Antonarakis S.E.,
 RA Antonarakis S.E., Krohn K.J.E.;
 RT "Mutation analyses of North American APS-1 patients.";
 RL Hum. Mutat. 13:69-74(1999).
 RN [9]
 RP VARIANT ARG-278.
 RX MEDLINE-98381937; PubMed-9717837;
 RA Scott H.S., Heino M., Peterson P., Miltaz L., Laijola M.D.,
 RA Bettele C., Cohen A., Serl M., Lerone M., Romeo G., Collin P.,
 RA Salo M., Metcalfe R., Weetman A., Papasavvas M.-P., Rossier C.,
 RA Nagamine K., Kudoh J., Shimizu N., Krohn K.J.E., Antonarakis S.E.;
 RT "Common mutations in autoimmune polyendocrinopathy-candidiasis-
 RT ectodermal dystrophy patients of different origins.";
 RL Mol. Endocrinol. 12:1112-1119(1998).
 RN [10]
 RP VARIANTS APECED LEU-15; MET-16; PRO-28; ARG-78; LEU-80;
 RX GLU-83; CYS-85; ARG-90; ARG-93; MET-301; TYR-311 AND GLN-326, AND
 RP VARIANT ARG-278.
 RX PubMed-11524731;
 RA Heino M., Peterson P., Kudoh J., Shimizu N., Antonarakis S.E.,
 RA Scott H.S., Krohn K.J.E.;
 RT "APECED mutations in the autoimmune regulator (AIRE) gene.";
 RL Hum. Mutat. 18:205-211(2001).
 CC -1- FUNCTION: Probable transcriptional regulator protein that binds to
 CC DNA as dimer and tetramer, but not as a monomer. Binds to G-
 CC tandems in an A/T-rich environment; the preferred motif is a
 CC tandem repeat of ATTGGTA combined with a TTAATTA-box. May be
 CC involved in immune regulation.
 CC -1- SUBUNIT: Homodimer and homotrimer. Interacts with CREBP.
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic; associated with
 CC tubular structures and in discrete nuclear dots resembling ND10
 CC nuclear bodies. May shuttle between nucleus and cytoplasm.
 CC -1- ALTERNATIVE PRODUCTS: At least 3 isoforms: 1/AIRE-1 (shown here),
 CC 2/AIRE-2 and 3/AIRE-3; may be produced by alternative
 CC splicing.
 CC -1- TISSUE SPECIFICITY: Widely expressed. Expressed at higher level in
 CC thymus (medullary epithelial cells and monocyte-dendritic cells),
 CC pancreas, adrenal cortex, and testis. Expressed at lower level in
 CC the spleen, fetal liver and lymph nodes. Isoforms 2 and 3 seem to
 CC be less frequently expressed than isoform 1, if at all.
 CC -1- DOMAIN: The L-X-X-L-L repeats may be implicated in binding to
 CC nuclear receptors.
 CC -1- DOMAIN: The HSR domain is required for localization on tubular
 CC structures (N-terminal part) and for homodimerization.
 CC -1- DOMAIN: Disruption of the first PHD domain has been shown to lead
 CC to reduced transcriptional activity and to localization of the

CC protein mainly in the cytoplasm in small granules.
 CC -1- PTM: Phosphorylated. Phosphorylation could trigger
 CC oligomerization.
 CC -1- DISEASE: Defects in AIRE are a cause of autoimmune poly-
 CC endocrinopathy candidiasis ectodermal dystrophy (APECED), also
 CC called autoimmune polyglandular syndrome type 1 (APS-1), an
 CC autosomal recessive disease characterized by: (1) autoimmune
 CC polyendocrinopathies: hypoparathyroidism, adrenocortical failure,
 CC IDDM, gonadal failure, hypothyroidism, pernicious anemia, and
 CC hepatitis; (2) chronic mucocutaneous candidiasis; (3) ectodermal
 CC dystrophies: vitiligo, alopecia, keratopathy, dystrophy of dental
 CC enamel, nails and tympanic membranes. In addition, a high
 CC proportion of patients develop squamous cell carcinoma of the oral
 CC mucosa. The disease is reported worldwide but is exceptionally
 CC prevalent among the Finnish population (incidence 1:25000) and
 CC the Iranian Jews (incidence 1:9000).
 CC -1- SIMILARITY: CONTAINS 1 HSR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
 CC -1- SIMILARITY: CONTAINS 1 SAND DOMAIN.
 CC -1- DATABASE: NAME-AIRE; NOTE-Article on AIRE and APECED;
 CC WWW="http://chr21.rz-berlin.mpg.de/APECED.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.slb.ch).
 CC -----
 CC EMBL; AB006682; BAA23988.1; -
 CC EMBL; AB006683; BAA23989.1; -
 CC EMBL; AB006684; BAA23990.1; -
 CC EMBL; AB006684; BAA23991.1; -
 CC EMBL; AB006684; BAA23992.1; -
 CC EMBL; AB006685; BAA23993.1; -
 CC EMBL; Z97990; CAB10790.1; -
 CC EMBL; AJ009610; CA008759.1; -
 CC EMBL; AF001754; BAA95560.1; -
 CC Genbank; HGNC:360; AIRE.
 CC MIM: 240300; -
 CC InterPro: IPR000770; SAND_domain.
 CC InterPro: IPR004865; Sp100.
 CC InterPro: IPR001965; ZnF_PHD.
 CC Pfam: PF00628; PHD; 2.
 CC Pfam: PF01342; SAND; 1.
 CC SMART: SM00249; PHD; 2.
 CC SMART: SM00258; SAND; 1.
 CC PROSITE: PS00864; SAND; 1.
 CC PROSITE: PS01359; ZF_PHD_1; 2.
 CC PROSITE: PS0016; ZF_PHD_2; 1.
 CC KW Nuclear protein; DNA-binding; Transcription regulation; Zinc-finger;
 CC Repeat; Phosphorylation; Alternative splicing; Polymorphism;
 CC Disease mutation.
 CC KW
 CC DOMAIN 1 106
 CC DOMAIN 181 280
 CC ZN_FING 296 343
 CC ZN_FING 434 475
 CC DOMAIN 113 133
 CC DOMAIN 7 520
 CC REPEAT 7 11
 CC REPEAT 63 67
 CC REPEAT 414 418
 CC REPEAT 516 520
 CC VANSPLIC 1 292
 CC
 CC HSR.
 CC SAND.
 CC PHD-TYPE 1.
 CC PHD-TYPE 2.
 CC NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 CC 4 X 5 AA REPEATS OF L-X-X-L-L.
 CC 1.
 CC 2.
 CC 3.
 CC 4.
 CC MATDAALRLRLRLHRTETAVAVDAFPLLLHALADHDVVED
 CC KFOETLHKEKGCQAFHALSWLTDSTAIIDFVRYLF
 CC KDYNLERGRQPIIDSPFKVDLSQPKRGKPPVAVYALF
 CC PPRPLPTKRSKSEARAAPALPFGTASVPGSOLKAPPK
 CC KPESASQOORPLDGLGIGTOMASVORAVAMSGDPCARGA
 CC VEGTILIOVFSSGSKKCIQVGGERTTSKEDSSGKAKA
 CC RSSSGPKLVAKAGQAGAPGGEARLQDQGSVPPALPS
 CC DPQLH -> MMLVYSSGAPGTQGPARNRVFPPIGMAPGVC

Query Match 100.0%; Score 2902; DB 1; Length 545;
 Best Local Similarity 100.0%; Pred. No. 3.3e-152;
 Matches 545; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATDAALRLRLHRTETVAVDSAPFLALADHDVVPEDKQETLHLKEBSCPOAFH 60
 |||||||
 DB 1 MATDAALRLRLHRTETVAVDSAPFLALADHDVVPEDKQETLHLKEBSCPOAFH 60
 |||||||

QY 61 ALLSWLLTQDSTAILDFWHLFEKDYNIERYGRLOPLIDSPFKVYDSQPKRGKRPVVK 120
 |||||||
 DB 61 ALLSWLLTQDSTAILDFWHLFEKDYNIERYGRLOPLIDSPFKVYDSQPKRGKRPVVK 120
 |||||||

QY 121 ALVPPRLPTKKASEEARAAAPALTPRGTA SPGSOLKAPKPKKRESSAEQOQLPLGNG 180
 |||||||
 DB 121 ALVPPRLPTKKASEEARAAAPALTPRGTA SPGSOLKAPKPKKRESSAEQOQLPLGNG 180
 |||||||

QY 181 IOTMSAVQRAVAMSSGDVPGARGAVEGILIOVFESGSKCICQYGEFTYPSKREDSG 240
 |||||||
 DB 181 IOTMSAVQRAVAMSSGDVPGARGAVEGILIOVFESGSKCICQYGEFTYPSKREDSG 240
 |||||||

QY 241 SGKNKARSSGPKPLVRAKAGAGAAPGGEGARLGOOGSVAPALPALSDDPOLHOKNDECA 300
 |||||||
 DB 241 SGKNKARSSGPKPLVRAKAGAGAAPGGEGARLGOOGSVAPALPALSDDPOLHOKNDECA 300
 |||||||

QY 301 VCRGGELICDGCPRAFHLACLSPLREIPSGTWRCSSCLQATVOEVOVPRAEPRPOEP 360
 |||||||
 DB 301 VCRGGELICDGCPRAFHLACLSPLREIPSGTWRCSSCLQATVOEVOVPRAEPRPOEP 360
 |||||||

QY 361 PVETPLPPLGRLSAGEVGPPEGLAGMDTLYVKHLRPPSAAPLPGLDSSALHPLLCV 420
 |||||||
 DB 361 PVETPLPPLGRLSAGEVGPPEGLAGMDTLYVKHLRPPSAAPLPGLDSSALHPLLCV 420
 |||||||

QY 421 GPEGOQNLAPGARGCVCGDGTDLVLRCTCAAFHNRCHPAGTSRPGTGLRCRSCGDVT 480
 |||||||
 DB 421 GPEGOQNLAPGARGCVCGDGTDLVLRCTCAAFHNRCHPAGTSRPGTGLRCRSCGDVT 480
 |||||||

QY 481 PAPVGVLAAPSPARLAPPAKDDTASHEPALHRDLESLSEHTEDGILQMAIQSMARPA 540
 |||||||
 DB 481 PAPVGVLAAPSPARLAPPAKDDTASHEPALHRDLESLSEHTEDGILQMAIQSMARPA 540
 |||||||

QY 541 APFPS 545
 |||||
 DB 541 APFPS 545

RESULT 2
 AIRE_MOUSE
 ID AIRE_MOUSE STANDARD; PRT: 552 AA.
 AC 0920B3; 09JUX0; 09JLM1; 09JLM2; 09JLM3; 09JLM4; 09JLM5; 09JLM6;
 AC 09JLM7; 09JLM8; 09JLM9; 09JLM0;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Autoimmune regulator (Autoimmune polyendocrinopathy candidiasis
 ectodermal dystrophy protein homolog) (APECED protein homolog).
 GN AIRE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1A).
 RA Shi J.-D., Wang C.-Y., Marron M.P., Ruan Q.-G., Huang Y.Q.,
 RT "Chromosomal localization and complete genomic sequence of the murine
 RT autoimmune regulator gene (Aire).";
 RL Autoimmunity 31:47-53(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1A).
 RC STRAIN=129/SV;
 RX MEDLINE=99160890; PubMed=10049735;

RA Miltaz L., Rossier C., Helmo M., Petersen P., Krohn K.J.E., Gos A.,
 RA Morris M.A., Kudoh J., Shimizu N., Antonarakis S.E., Scott H.S.,
 RT "Isolation and characterization of the mouse Aire gene.";
 RL Biochem. Biophys. Res. Commun. 255:483-490(1999).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1A).
 RC STRAIN=C57BL/6; TISSUE=Thymus;
 RX MEDLINE=99168902; PubMed=10049587;
 RA Wang C.-Y., Shi J.-D., Davoodi-Semlromi A., Shi J.-X.,
 RT "Cloning of Aire, the mouse homologue of the autoimmune regulator
 RT (AIRE) gene responsible for autoimmune polyglandular syndrome type 1
 RT (ASPI).";
 RL Genomics 55:322-326(1999).
 RN [4]
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
 RC STRAIN=129;
 RX PubMed=10022980;
 RA Blechschmidt K., Schweiger M., Wertz K., Poulsen R.,
 RA Christensen H.-M., Rosenthal A., Lehrach H., Yaspo M.-L.,
 RT "The mouse Aire gene: comparative genomic sequencing, gene
 RT organization, and expression.";
 RL Genome Res. 9:158-166(1999).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 1A; 1B; 1C; 1D; 2A; 2B; 2C; 2D; 3A; 3B;
 RP 3C AND 3D).
 RC STRAIN=C57BL/6, NOD, and STL;
 RX PubMed=10550218;
 RA Ruan Q.-G., Wang C.-Y., Shi J.-D., Shi J.-X.,
 RT "Expression and alternative splicing of the mouse autoimmune regulator
 RT gene (Aire).";
 RL J. Autoimmun. 13:307-313(1999).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM 1A), AND SUBCELLULAR LOCATION.
 RC TISSUE=Kidney, and Thymus;
 RX PubMed=11156688;
 RA Halonen M., Peltto-Huikko M., Eskelin P., Peltonen L., Ullmanen I.,
 RA Kolmer M.,
 RT "Subcellular location and expression pattern of autoimmune regulator
 RT (Aire), the mouse orthologue for human gene defective in autoimmune
 RT polyendocrinopathy candidiasis ectodermal dystrophy (APECED).";
 RL J. Histochem. Cytochem. 49:197-208(2001).
 RN [7]
 RP SUBUNIT STRUCTURE, AND PHOSPHORYLATION.
 RC PubMed=11533054;
 RX Kumar P.G., Laloraya M., Wang C.-Y., Ruan Q.-G., Davoodi-Semlromi A.,
 RA Kao K.-J., Shi J.-X.,
 RT "The autoimmune regulator (AIRE) is a DNA-binding protein.";
 RL J. Biol. Chem. 276:41357-41364(2001).
 CC -1- FUNCTION: Probable transcriptional regulator protein that binds to
 CC DNA as dimer and tetramer, but not as a monomer. Binds to G-
 CC tandem repeat in an A/T-rich environment; the preferred motif is a
 CC involved in immune regulation (By similarity).
 CC -1- SUBUNIT: Homodimers and homotetramers.
 CC -1- SUBCELLULAR LOCATION: Predominantly nuclear but also cytoplasmic.
 CC Found in nuclear body-like structures and in a filamentous
 CC vimentin-like pattern.
 CC -1- ALTERNATIVE PRODUCTS: At least 12 isoforms; 1a (shown here), 1b,
 CC 1c, 1d, 2a, 2b, 2c, 2d, 3a, 3b, 3c and 3d; are produced by
 CC alternative splicing. Isoforms 3a to 3d are probably
 CC nonfunctional.
 CC -1- TISSUE SPECIFICITY: Expression may be restricted to a small number
 CC of scattered cells in most tissues. Highly expressed in a few
 CC cells in the medulla of the thymus. Detected at very low levels in
 CC thymus, lymph node, liver, brain, ovary, lung, testis, kidney,
 CC heart, spleen, bone marrow, skeletal muscle and adrenal gland.
 CC Isoforms 1a to 1d predominate, isoforms 2a to 2d are intermediate
 CC and isoforms 3a to 3d are expressed at extremely low levels.
 CC -1- DOMAIN: The L-X-X-L-L repeats may be implicated in binding to
 CC nuclear receptors.
 CC -1- DOMAIN: The N-terminal HSR domain is required for localization
 CC on tubular structures (By similarity).
 CC -1- PTM: Phosphorylated.


```

CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DISEASE: ONE OF THE MAIN ANTIGENS REACTING WITH ANTI-MI-2 POSITIVE
CC SERA OF DERMATOMYOSITIS.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CHROMO DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF06515; AAB87383.1; -.
DR EMBL: 008379; AAC50228.1; -.
DR Genew: HGNC:1918; CHD3.
DR MIM: 602120; -.
DR InterPro: IPR000953; Chromo.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002464; DEAH_box.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR InterPro: IPR001841; Znf_PHD.
DR Pfam: PF00176; SNF2_N; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00385; chromo; 1.
DR Pfam: PF00628; PHD; 2.
DR SMART: SM00298; CHROMO; 2.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR SMART: SM00249; PHD; 2.
DR SMART: SM00184; RING; 2.
DR PROSITE: PS00598; CHROMO_2; 1.
DR PROSITE: PS0013; CHROMO_2; 2.
DR PROSITE: PS00690; DEAH-ATP-HELICASE; 1.
DR PROSITE: PS01359; ZF_PHD_1; 2.
DR PROSITE: PS00016; ZF_PHD_2; 2.
KW Chromatin regulator; DNA-binding; ATP-binding; Helicase;
KW Nuclear protein; Repeat; Transcription regulation; Activator; Antigen;
KW Zinc-finger.
FT ZN_FING 379 426 PHD-TYPE 1.
FT ZN_FING 456 503 PHD-TYPE 2.
FT DOMAIN 494 594 CHROMO 1.
FT DOMAIN 631 673 CHROMO 2.
FT NP_BIND 761 768 ATP (POTENTIAL).
FT SITE 883 886 DEAH_BOX.
FT DOMAIN 206 221 POLY-ALA.
FT DOMAIN 243 246 POLY-PRO.
FT DOMAIN 355 358 POLY-LYS.
FT DOMAIN 434 446 POLY-GLU.
FT DOMAIN 697 703 POLY-LYS.
FT CONFLICT 121 126 GEGDGG->PHFQOK (IN REF. 2).
FT CONFLICT 309 312 MISSING (IN REF. 2).
FT CONFLICT 653 653 W->G (IN REF. 2).
SQ SEQUENCE 1944 AA; 220691 MW; 0712E6F1D5FD7D335 CRC64;

Query Match 9.3%; Score 270.5; DB 1; Length 1944;
Best Local Similarity 23.2%; Pred. No. 1.1e-07;
Matches 102; Conservative 46; Mismatches 154; Indels 137; Gaps 17;

```

```

DB 310 VFQSDGEPPEAESDLDGSHASGRPDGPFVTKKLRGRGRKKKKYLQ----CPAV 365
QY 283 LALPSDPOLHOKNEDECAVCPRDGEILCCDGPRAFLACLSPLEIPSGTWRSCSLQ 342
DB 366 AGESEVQYETFDHOCYCEVCOGGELLCDTQRAHLVCLDPELDRAFGKSCPCXK 425
QY 343 ATVOEVOGRAEPRPOEPPVETPLPPGLRSAGEEVNPGPEPLAGMDTLVYKHLPAAPS 402
DB 426 EGVQWEAKEEEVEEVEE-----GEE-EGEKEE-----DHMEY----- 458
QY 403 AARPLGDDSLHPLLCVGPBEGQNLAPARGCYGCDGTDLVRLCTHCAAFHNRCHFPAG 462
DB 459 -----CRVCKDGGELLCCDCAISSYHTHCLMPPL 487
QY 463 TSPRGTLRCRSCGSDVTPAPVEGYL-----APSPARLAPAKODTASHEPALH 512
DB 488 PDINGEWLRCRCCTCPVLKGRQKILHWRGEPVAVPAPQADGNP---DVPPRPLQG 544
QY 513 RDLLESLSEHTFDGIILQW 531
DB 545 RSEREFVVK---WVGLSYW 560

RESULT 4
CHD4_HUMAN STANDARD; PRT; 1912 AA.
ID CHD4_HUMAN
AC Q14839;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromodomain helicase-DNA-binding protein 4 (CHD-4) (MI-2 autoantigen
DE 218 kDa protein) (MI2-beta).
GN CHD4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=6017437; Pubmed=7575689;
RA Seelig H.P., Moosbrugger T., Ehrfeld H., Fink T., Renz M., Genth E.;
RT "The major dermatomyositis specific Mi-2 autoantigen is a presumed
RT helicase involved in transcriptional activation."
RL Arthritis Rheum. 38:1389-1399(1995).
CC -1- FUNCTION: POSSIBLE TRANSCRIPTION ACTIVATOR (BY SIMILARITY).
CC -1- SUBUNIT: Central component of the nucleosome remodelling and
CC histone deacetylase (NURD) complex.
CC -1- DISEASE: ONE OF THE MAIN ANTIGENS REACTING WITH ANTI-MI-2 POSITIVE
CC SERA OF DERMATOMYOSITIS.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 2 CHROMO DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X86691; CAA60384.1; -.
DR Genew: HGNC:1919; CHD4.
DR MIM: 603277; -.
DR InterPro: IPR000953; Chromo.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002464; DEAH_box.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR InterPro: IPR001965; Znf_PHD.
DR InterPro: IPR001841; Znf_PHD.
DR Pfam: PF00176; SNF2_N; 1.

```

DR Pfam: PF00271; helicase-C; 1.
 DR Pfam: PF00385; chromo; 1.
 DR Pfam: PF00628; PHD; 2.
 DR SMART: SM00298; CHROMO; 2.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR SMART: SM00249; PHD; 2.
 DR SMART: SM00184; RING; 2.
 DR PROSITE: PS00598; CHROMO_1; 2.
 DR PROSITE: PS50013; CHROMO_2; 2.
 DR PROSITE: PS00690; DEAH_ATP_HELICASE; 1.
 DR PROSITE: PS01359; ZF_PHD_1; 2.
 DR PROSITE: PS50016; ZF_PHD_2; 2.
 DR Chromatin regulator; DNA-binding; ATP-binding; Helicase;
 KW Nuclear protein; Repeat; Transcription regulation; Activator; Antigen;
 FT ZN_FING 370 417 PHD-TYPE 1.
 FT ZN_FING 449 496 PHD-TYPE 2.
 FT DOMAIN 494 594 CHROMO 1.
 FT DOMAIN 622 697 CHROMO 2.
 FT NP_BIND 751 758 ATP (POTENTIAL).
 FT SITE 873 876 DEAH_BOX.
 FT DOMAIN 50 59 POLY-LYS.
 FT DOMAIN 94 98 POLY-GLU.
 FT DOMAIN 114 119 POLY-LYS.
 FT DOMAIN 134 138 POLY-GLU.
 FT DOMAIN 139 144 POLY-ASP.
 FT DOMAIN 227 235 POLY-ALA.
 FT DOMAIN 248 252 POLY-PRO.
 FT DOMAIN 350 354 POLY-LYS.
 FT DOMAIN 1052 1055 POLY-LEU.
 FT DOMAIN 1294 1301 POLY-GLU.
 FT DOMAIN 1665 1668 POLY-GLU.
 SQ SEQUENCE 1912 AA; 217989 MW; 80A73300E80166CE CRC64;

Query Match 8.38; Score 241; DB 1; Length 1912;
 Best Local Similarity 23.18; Pred. No. 4.4e-06;
 Matches 97; Conservative 51; Mismatches 172; Indels 100; Gaps 14;

QY 129 PTKKASEEAAAPALTPRGTSAPGSQLAKPKPKRESSAEQORLPNGISQIMASAV 188
 Db 217 PFKSSGASVAAAAAAYAVESWYATREV-APPPEVETPIRKAKTRGKGPMA----- 270
 QY 189 QRAVAMSSGDVPGARVAGVEGILIOVEPSG-----SKKCIQVGEFYPSKPED----- 238
 Db 271 -RRPKSGSPRPDAKKPKPKVAPLKIKIGFGSKRKSSSSDDDLVESDPDASINSY 329
 QY 239 SGSGKKARSSSGPKPLVRAKGAAGAPGGEARLGQGSVPAPLALPSDQLHOKNEDE 298
 Db 330 SVSGSGSRSSRSRKKLRITK-----KKRKGEE-EVTAVDGYETD-----HODY 372
 QY 299 CAVCRDGEILICDCGPRAFILACTSPPLREIPSGTWRCSSCLQATVQEVOPRAEPRQ 358
 Db 373 CEVCQCGGEIILCTCRAITHWCLDPMKAPGKMSCPCEK---EGIQWEAKED--- 426
 QY 359 EPPVETPLPGRLSAGEEVGPPPEPLAGMDTLVYKHLPAAPSAAPLGLDSSALHPL 418
 Db 427 -----NSEGEILIEVVGDLDEED-----HHMEF----- 451
 QY 419 CVGPEGOQNLAPGARCGVGDGDVLRCTHCAAFHWRCHFPAGTSRPGTLRCRSCGD 478
 Db 452 -----CVCKCKDGGELLCDDTCPSYHICLNPLPEIPNGEWLCPRTCP 496
 QY 479 VTPAPEGVV-----APSPARL-APPAKDDTASHPEPALHRDLESLSEHTDGILOW 531
 Db 497 ALKGKVKILIKWKGPPSPFPVPRPPDADPNPSPKPLEGRBPRQPFVK---WQGMKYW 553

RESULT 5
 SP11_HUMAN
 ID SP11_HUMAN STANDARD: PRT; 689 AA.
 AC Q9H58; Q9H58; Q9H58; Q14976; Q14977;
 DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sp110 nuclear body protein (Speckled 110 kDa) (Transcriptional
 DE coactivator Sp110) (interferon-induced protein 41/75).
 GN SP110.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4), AND PHOSPHORYLATION (ISOFORM
 RP 2).
 RC TISSUE=Lymphoma;
 RA MEDLINE=94043285; PubMed=7693701;
 RT Kaderoff S., Gewert D.R., Galabru J., Hovanessian A.G., Meurs E.F.;
 RT "Molecular cloning of two new interferon-induced, highly related
 RT nuclear phosphoproteins.";
 RL J. Biol. Chem. 268:24432-24441(1993).
 [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND SUBCELLULAR LOCATION.
 RC TISSUE=Spleen;
 RX MEDLINE=20372745; PubMed=10913195;
 RA Bloch D.B., Nakajima A., Gulick T., Chalche J.-D., Orth D.,
 RA de la Monte S.M., Bloch K.D.;
 RT "Sp110 localizes to the PML-Sp100 nuclear body and may function as a
 RT nuclear hormone receptor transcriptional coactivator.";
 RL Mol. Cell. Biol. 20:6138-6146(2000).
 [3]
 RN SEQUENCE FROM N.A. (ISOFORM 5).
 RP TISSUE=Lymph;
 RC Strausberg R.;
 RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Transcription factor. May be a nuclear hormone receptor
 CC coactivator. Enhances transcription of genes with retinoic acid
 CC response elements (RARE).
 CC -1- SUBCELLULAR LOCATION: Nuclear; found in the nuclear body.
 CC -1- ALTERNATIVE PRODUCTS: 5 isoforms; 1 (shown here), 2/FT15/75,
 CC 3/Sp110b, 4/FT14/41 and 5; are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Highly expressed in peripheral blood
 CC leukocytes and spleen. Detected at intermediate levels in thymus,
 CC prostate, testis, ovary, small intestine and colon, and at low
 CC levels in heart, brain, placenta, lung, liver, skeletal muscle,
 CC kidney and pancreas.
 CC -1- INDUCTION: By interferon gamma and by all-trans retinoic acid.
 CC -1- PTM: Phosphorylated (isoform 2).
 CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 HSR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC -1- SIMILARITY: CONTAINS 1 SAND DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L22342; AAA18806.1; -
 CC EMBL: L22343; AAD13402.1; -
 CC EMBL: AF280094; AAF99318.1; -
 CC EMBL: AF280095; AAG09826.1; -
 CC EMBL: BC019059; AAH19059.1; -
 CC Genew: HGNC:5401; SP110.
 CC MIM: 604457; -
 CC InterPro: IPR001487; Bromodomain.
 CC InterPro: IPR00770; SAND_domain.
 CC InterPro: IPR004865; SP100.
 CC InterPro: IPR001965; Znf_PHD.
 CC Pfam: PF00628; PHD; 1.
 CC Pfam: PF01342; SAND; 1.
 CC Pfam: PF03172; SP100; 1.
 CC SMART: SM00297; BROMO; 1.

DR SMART: SM00249; PHD: 1.
 DR SMART: SM00258; SAND: 1.
 DR PROSITE: PS50864; SAND: 1.
 DR PROSITE: PS50016; ZF_PHD_2; 1.
 DR PROSITE: PS50063; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE: PS50014; BROMODOMAIN_2; FALSE_NEG.
 DR Nuclear protein; DNA-binding; Trans-acting factor; Zinc-finger;
 KW Phosphorylation; Bromodomain; Alternative splicing.
 FT DOMAIN 6 109 HSR.
 FT DOMAIN 281 294 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 428 444 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 454 535 SAND.
 FT DOMAIN 525 529 NUCLEAR HORMONE RECEPTOR INTERACTION
 (POTENTIAL).
 FT ZN_FING 534 580 PHD-TYPE.
 FT DOMAIN 581 676 BROMODOMAIN.
 FT VARSPPLIC 1 251 MISSING (IN ISOFORM 4).
 FT VARSPPLIC 1 203 MISSING (IN ISOFORM 2).
 FT VARSPPLIC 141 143 HHR -> PPO (IN ISOFORM 5).
 FT VARSPPLIC 252 275 IRDNPEDPEEPQEVSTPSDK -> MASSGVKNTPRMR
 RKAPGREKE (IN ISOFORM 4).
 FT VARSPPLIC 300 349 MISSING (IN ISOFORM 4).
 FT VARSPPLIC 300 303 GTAS -> AL (IN ISOFORM 5).
 FT VARSPPLIC 531 549 RKNSECEVCOCGQLCC -> SCGLCPPTINKRELNS
 K (IN ISOFORM 4 AND ISOFORM 3).
 FT VARSPPLIC 531 539 RKNSECEV -> SGLCSVLOE (IN ISOFORM 3).
 FT VARSPPLIC 540 689 MISSING (IN ISOFORM 3).
 FT VARSPPLIC 550 689 MISSING (IN ISOFORM 4 AND ISOFORM 5).
 FT VARSPPLIC 606 611 IRDGE -> NVSSSS (IN ISOFORM 2).
 FT VARSPPLIC 612 689 MISSING (IN ISOFORM 2).
 FT CONFLICT 167 167 T -> D (IN REF. 3).
 FT CONFLICT 367 367 T -> M (IN REF. 1); AAAL8806).
 FT CONFLICT 464 464 L -> S (IN REF. 1; AAAL8806).
 FT CONFLICT 523 523 M -> T (IN REF. 2).
 SO SEQUENCE 689 AA; 78504 MM; EBBE023101C2E59B CRC64;

Query Match 7.9%; Score 228; DB 1; Length 689;
 Best Local Similarity 10.5%; Pred. No. 8.2e-06;
 Matches 119; Conservative 44; Mismatches 138; Indels 308; Gaps 20;

2 AYDAALRLRLHREIIVAVDSAPFLHALADHDVVPEDKFOETLHKKEG----- 55
 7 AMEALLOHFNHOKIGIAIHKPPFFEGGLDNIITKRMWESL-----EACRNLIPIV 61
 56 POAFHALLSWLTDSTALDFWRYLFKDNLERYGRLOPIIDSPKVDLSOPRKGR-- 113
 62 SRVYHNILTOI---ERTFNLILVLFESQINIREYDNLVTYHSF-KRVGASVERQGRDT 117
 114 -----KP-----PAVPKALVP----- 124
 118 PILLEAPTLAEGSSLHRLRLALPHKRPPOSCPCAPRVSEPTSSQSTELISESPSPS 177
 125 .PRPLP-----TKRKASEEARAARPAALT----- 147
 178 DPVLPRLALIOBGRSTVTKDKLTKMAEEDSEEMPSLITVQVASDMLPIQIRKED 237
 148 ----- 147
 238 POEMPHSLGSMPEIRDNSPEDEPEQEVSTPSDKKKKKRRCIMSTPKRRHKKKSL 297
 148 PRGASP--GSQLAK-----PKKPRESSA----- 170
 298 PRGTAASHHGLOKLRVDYQOKDDSTCNSTVETRAOKARTECARKSRSEELIDTSE 357
 171 -----EORLPL-----GNGIO-----TMSASYORA-- 191
 358 MNEGRSKQKTPSTPRRYVQGAASPGHGIOEKLOVDKVTORKDSTWNSVMMARVQAPART 417
 192 -----VAMSSGVPGARGAVEGLILO 212
 418 KCAKRSRKKEKKKEDICSSSKRRFQKNIHRGKPKSPDVFDFCSKLPVCGEAGKILYK 477
 213 QVESGSGKCLIQ-VGGEFTYTSKFEDSDSGKNKARSSSGPKPLVLRKAGQGAAPGGEA 271

Db 478 KKKHSSSVKCIKNEEDGTWTFPEFEYEGKGRN-----AKMKRNIKRG-----M 523
 Qy 272 RFGOGSVAPLALPSPDPOHKNEDCAVCARDGELICDGCRAFLACLSPLEIRIP 331
 Db 524 TLGE-----LLKRNSECEVCOCGQLCCGTCGRVFNEDCHIPPV-EAK 568
 Qy 332 SCTWRCCSC 340
 Db 569 RMLWCTFC 577

RESULT 6

TF1A_HUMAN STANDARD; PRT; 1050 AA.
 AC 015164; 095854;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Transcription Intermediary factor 1-alpha (TIF1-alpha).
 OS TF1A OR TF1.
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RC TISSUE=breast cancer;
 RX MEDLINE=97277352; PubMed=9115274;
 RA Therot S., Henriquet C., Rochefort H., Cavaillès V.;
 RT "Differential interaction of nuclear receptors with the putative human
 transcriptional coactivator hTIF1.";
 RL J. Biol. Chem. 272:12062-12068(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RX MEDLINE=99144725; PubMed=10022127;
 RA Venturini L., You J., Stadler M., Gallen R., Lallemand V.,
 RA Koken H.H., Mattei M.-G., Ganser A., Chambon P., Losson R.,
 RA De The H.;
 RT "TIF1gamma, a novel member of the transcriptional intermediary factor
 1 family.";
 RL Oncogene 18:1209-1217(1999).
 RN [3]
 RP SEQUENCE OF 477-510 (LONG ISOFORM).
 RC TISSUE=breast cancer;
 RA Cavaillès V.;
 RL Submitted (JAN-1999) to the SWISS-PROT data bank.
 CC -!- FUNCTION: INTERACTS SELECTIVELY IN VITRO WITH THE AR2-ACTIVATING
 DOMAIN OF THE ESTROGEN RECEPTORS. ASSOCIATION WITH DNA-BOUND
 ESTROGEN RECEPTORS REDUCES THE PRESENCE OF ESTRADIOL.
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -!- SIMILARITY: CONTAINS 2 B BOX-TYPE ZINC FINGERS.
 CC -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: AF009353; AAB63585.1; -;
 DR EMBL: AF119042; AAD17258.1; -;
 DR HSSP: P29590; 1BOR.
 DR TRANSFAC: T04945; -;
 DR Genew: HGNC:11812; TIF1.
 DR MIM: 603406; -;
 DR InterPro: IPR003649; Bbox_C.

DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR000315; Znf_Box.
 DR InterPro: IPR001965; Znf_PHD.
 DR InterPro: IPR001841; Znf_Ring.
 DR Pfam: PF00097; Zf-C3HC4. 1.
 DR Pfam: PF00439; bromodomain; 1.
 DR Pfam: PF00628; PHD; 1.
 DR Pfam: PF00643; Zf-B_Box; 2.
 DR PRINTS: PR01406; BBOXZNFINGER.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00502; BBC; 1.
 DR SMART: SM00336; BBOX; 2.
 DR SMART: SM00297; BBOX; 1.
 DR SMART: SM00249; PHD; 1.
 DR SMART: SM00184; RING; 1.
 DR PROSITE: PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 DR PROSITE: PS01359; ZF-PHD_1; 1.
 DR PROSITE: PS50016; ZF-PHD_2; 1.
 DR PROSITE: PS50018; ZF-RING_1; 1.
 DR PROSITE: PS50089; ZF-RING_2; 1.
 DR Transcription regulation; Repressor; DNA-binding; Bromodomain;
 DR Zinc-finger; Alternative splicing; Nuclear protein; Coiled coil;
 Repeat.
 KW Repeat.
 FT DOMAIN 9 15 POLY-ALA.
 FT ZN_FING 56 82 RING-TYPE.
 FT ZN_FING 158 211 B BOX-TYPE 1.
 FT ZN_FING 218 259 B BOX-TYPE 2.
 FT DOMAIN 289 359 COILED COIL (POTENTIAL).
 FT DOMAIN 344 347 POLY-GLN.
 FT DOMAIN 754 779 NUCLEAR RECEPTOR BINDING SITE (NRBS).
 FT ZN_FING 826 873 PHD-TYPE.
 FT DOMAIN 891 907 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 932 987 BROMODOMAIN.
 FT VARSPLIC 477 510 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 14 20 AASAAS -> RLGAP (IN REF. 1).
 FT CONFLICT 24 28 SAAPS -> RGG (IN REF. 1).
 FT CONFLICT 109 114 GSPVSG -> RGG (IN REF. 1).
 FT CONFLICT 350 350 D -> N (IN REF. 1).
 FT CONFLICT 600 600 M -> I (IN REF. 1).
 FT CONFLICT 608 608 A -> R (IN REF. 1).
 FT CONFLICT 967 967
 SO SEQUENCE 1050 AA; 116831 MW; D341E8022AAC67E CRC64;

Query Match 7.8%; Score 227; DB 1; Length 1050;
 Best Local Similarity 24.6%; Pred. No. 1.4e-05;
 Matches 77; Conservative 46; Mismatches 104; Indels 86; Gaps 10;

71 SRAIDFVAVLEKRDYLERGLQPIIDSEPKVDLSQPRKGRKPPAVPALVPPRLPT 130
 631 STIMLD--IVRKDTNID-HGQPRPPSNRTVQSPNSVPSGLAGPYTMTSVHPPIRSPS 687
 131 KRKASEEAAAPALTPRGTA--SGLAKPKK-----PSSA 170
 688 --ASSVSGSSGSSSKPKGADSTHKVPVMEPIRIKQENSPPENYDEPVIVKQESD 745
 171 EQQLPLGNGIQTMSASVQRAVAMS-----GDVPGRGVDEGLILQOVESGSKK 222
 746 EESRPNANYPRIISLTLSLSSOSTSREYVLRSDAPSTGQPL----- 792
 223 CIYVGGEYTPSKFSDSGSKKARSSSGPKPLVRAKGAQGAAPGGEARLGGQSVAP 282
 793 -----HODNSSNGKSEMDPSOKSPL-----HYGEIR----- 819
 283 LALPSDPOLHOKNDECAVCHDGGELICCDGPPRAFHACLSPPLEIPSGTWKSCSLQ 342
 820 --KDDP-----NEDMCVAVCGGELLCCCKPKVFKHSCHVPTLTNPSGEMICFCMD 872
 343 ATVOEVOPRAPRP 355
 873 LSKREVEYDCDAP 885

RESULT 7
 ID CHDM_DROME STANDARD; PRT; 1982 AA.
 AC 097159; Q9VW50;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Chromodomain helicase-DNA-binding protein M1-2 homolog (Dml-2).
 GN M1-2 OR CG8103.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND MUTAGENESIS OF GLY-737.
 RX MEDLINE=99055400; PubMed=9836641;
 RA Kahle J., Beuchle D., Treuhelt S., Christen B., Kennison J.A.,
 RA Blenz M., Muller J.;
 RT "Dml-2, a hunchback-interacting protein that functions in Polycomb
 repression";
 RL Science 282:1897-1900(1998).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherier S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Bouchard M.R., Bouck J., Brockstein P., Brotler P.,
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Clawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunn B.C., Dunn P.,
 RA Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jajaeli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Laake P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liao X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sledge-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Sylvestre R., Tector C., Turner R., Venter M., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weisslock G.M., Weisslock J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: VITAL ROLE IN DEVELOPMENT. PROTEIN BINDS TO A PORTION OF
 HUNCHBACK (HB) PROTEIN THAT IS CRITICAL FOR REPRESSION OF BITHORAX
 COMPLEX (BXC) GENES. MAY ALSO FUNCTION IN POLYCOMB GROUP (PCG)
 REPRESSION OF HOX GENES.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 G.M. FAMILY.
 CC -!- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.

CC -1 CAUTION: Ref.2 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF119716; AAD17276.1; -
 DR EMBL: AE003515; AAF49099.1; ALT_SEQ.
 DR FlyBase: FBgn0013591; M1-2.
 DR InterPro: IPR000953; Chromo.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002464; DEAD_BOX.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR000330; SNF2_N.
 DR InterPro: IPR001965; Znf_PHD.
 DR InterPro: IPR001841; Znf_PHD.
 DR Pfam: PF00176; SNF2_N; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00385; Chromo; 1.
 DR Pfam: PF00628; PHD; 2.
 DR SMART: SM00298; CHROMO; 2.
 DR SMART: SM00487; DEAD; 1.
 DR SMART: SM00490; HELIC; 1.
 DR SMART: SM00249; PHD; 2.
 DR SMART: SM00184; RING; 2.
 DR PROSITE: PS00598; CHROMO_2; 2.
 DR PROSITE: PS00013; CHROMO_2; 2.
 DR PROSITE: PS00690; DEAD_BOX; 1.
 DR PROSITE: PS01359; ZF_PHD_1; 2.
 DR PROSITE: PS01359; ZF_PHD_2; 2.
 DR DNA-binding: ATP-binding; Helicase; Nuclear protein; Repeat;
 DR Transcription regulation; Repressor; Zinc-finger.
 KM ZN_FING 377 424 PHD-TYPE 1.
 FT ZN_FING 437 484 PHD-TYPE 2.
 FT ZN_FING 488 566 CHROMO 1.
 FT DOMAIN 612 673 CHROMO 2.
 FT NP_BIND 755 762 ATP (POTENTIAL).
 FT SITE 875 878 DEAD_BOX.
 FT DOMAIN 13 16 POLY-GLU.
 FT DOMAIN 70 76 POLY-LYS.
 FT DOMAIN 239 248 POLY-GLU.
 FT DOMAIN 1279 1287 POLY-GLU.
 FT DOMAIN 1672 1677 POLY-ASP.
 FT MUTAGEN 737 737 G->D: IN ALLELE M1-2-5; LARVAL LETHAL.
 FT CONFLICT 101 101 G->A (IN REF. 1).
 SQ SEQUENCE 1982 AA; 224199 MW; ED8E256D1AD0AC2P CRC64;
 Query Match 7.8%; Score 227; DB 1; Length 1982;
 Best Local Similarity 20.2%; Pred. No. 2,7e-05;
 Matches 77; Conservative 39; Mismatches 93; Indels 172; Gaps 11;

DB 417 SCFHC-----EADGAAEEEDDEHOF----- 439
 QY 396 HLPAPSAAPLPGIDSSALHPLLCVGPQQNLAPGARGCGDGTDLRCHCAAFHW 455
 DB 440 -----CRVCKDGGELLCDCSPSAHYR 461
 QY 456 RCHFPACTSRPGTGLRCRSCS 476
 DB 462 FCUNPLDRTIPDGDWRCRCS 482
 RESULT 8
 TFLA_MOUSE
 ID TFLA_MOUSE STANDARD; PRT; 1051 AA.
 AC 064127; 064126;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcription intermediary factor 1-alpha (TIF1-alpha).
 GN TIF1A OR TIF1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Carcinoma;
 RX MEDLINE=95262642; Pubmed=7744009;
 RA Le Douarin B., Zechel C., Garnier J.-M., Lutz Y., Tora L., Pierrat B.,
 RA Heery D., Gronemeyer H., Chambon P., Losson R.;
 RT "The N-terminal part of TIF1, a putative mediator of the ligand-
 RT dependent activation function (AF-2) of nuclear receptors, is fused to
 RT B-raf in the oncogenic protein T18";
 RL EMBO J. 14:2020-2033(1995).
 CC -1 FUNCTION: INTERACTS SELECTIVELY IN VITRO WITH THE AF2-ACTIVATING
 CC DOMAIN OF THE ESTROGEN RECEPTORS. ASSOCIATION WITH DNA-BOUND
 CC ESTROGEN RECEPTORS REQUIRES THE PRESENCE OF ESTRADIOL (BY
 CC SIMILARITY).
 CC -1 SUBCELLULAR LOCATION: Nuclear.
 CC -1 ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM. ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1 TISSUE SPECIFICITY: DETECTED IN ALL ADULT TISSUES. WITH THE
 CC HIGHEST EXPRESSION LEVEL IN TESTIS.
 CC -1 DISEASE: PARTICIPATES IN A CHROMOSOMAL TRANSLOCATION THAT PRODUCES
 CC A TFLA-BRAF (T18) ONCOGENE ORIGINALLY ISOLATED FROM A FURFURAL-
 CC INDUCED HEPATOMA.
 CC -1 SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -1 SIMILARITY: CONTAINS 2 B BOX-TYPE ZINC FINGERS.
 CC -1 SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC -1 SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S78221; AAB34290.1; -
 DR EMBL: S78219; AAB34289.1; -
 DR HSSP: P29590; 1BOR.
 DR TRANSFAC: T02143; -
 DR TRANSFAC: T02215; -
 DR MGD: MGI:109275; Tif1a.
 DR InterPro: IPR003649; Bbox_C.
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR000315; Znf_Box.
 DR InterPro: IPR001965; Znf_PHD.
 DR InterPro: IPR001841; Znf_PHD.
 DR Pfam: PF00439; bromodomain; 1.
 DR Pfam: PF00439; bromodomain; 1.

DR Pfam: PF00628; PHD; 1.
 DR Pfam: PF00643; zf-B-box; 2.
 DR PRINTS: PR01406; BROXZNFINGR.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR SMART: SM00302; BBOX; 1.
 DR SMART: SM00336; BBOX; 2.
 DR SMART: SM00297; BROMO; 1.
 DR SMART: SM00249; PHD; 1.
 DR SMART: SM00184; RING; 1.
 DR PROSITE: PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE: PS0014; BROMODOMAIN_2; 1.
 DR PROSITE: PS0119; ZF-BBOX; 2.
 DR PROSITE: PS01359; ZF-PHD; 1.
 DR PROSITE: PS0016; ZF-PHD; 2; 1.
 DR PROSITE: PS00518; ZF-RING; 1.
 DR PROSITE: PS00089; ZF-RING; 2; 1.
 DR Transcription regulation; Repressor; DNA-binding; Bromodomain;
 KW Zinc-finger; Alternative splicing; Nuclear protein; Coiled coil;
 KW Repeat; Proto-oncogene; Chromosomal translocation.
 FT DOMAIN 8 15
 FT ZN_FING 52 22 POLY-ALA.
 FT ZN_FING 158 211 RING-TYPE.
 FT ZN_FING 218 259 B BOX-TYPE 1.
 FT DOMAIN 289 359 B BOX-TYPE 2.
 FT DOMAIN 344 347 COILED COIL (POTENTIAL).
 FT DOMAIN 583 587 POLY-SER.
 FT DOMAIN 755 780 NUCLEAR RECEPTOR BINDING SITE (NRBS).
 FT ZN_FING 827 874 PHD-TYPE.
 FT DOMAIN 933 988 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT SITE 332 333 BREAKPOINT FOR TRANSLLOCATION TO FORM
 FT VARSPLIC MISSING (IN SHORT ISOFORM).
 FT SEQUENCE 1051 AA; 116656 MW; 610584C1C6885972 CRC64;

Query Match 7.5%; Score 217; DB 1; Length 1051;
 Best Local Similarity 27.2%; Pred. No. 5e-05;
 Matches 85; Conservative 46; Mismatches 122; Indels 60; Gaps 15;

QY 71 STALDEWRLVLFKDYNERGRLOPILDSFPKVDLSQ-----PKRRKPPAVKALVPP 126
 DB 606 SSPMIDLSAPVGSYSN-----LPSLP-DIDCSSTIMLDNARKDTGDHA---Qp 651
 QY 127 RLPTKRKASPEARAAAPALPLR-CTASPGQLAKPP-KKPESAEQRLPLNGIGIOTM 184
 DB 652 RPSNKTQV-----SNSSVSPFGLAGPYMTSVHPPIRNSPSSVSGSSG----- 699
 QY 185 SASQRAVAMSSGDVPGARGAVEGILIQVEFGSGSKK-----CIOVGGEFTYTSKFEDS 239
 DB 700 -SSSKPAGADSTHKVPV--MLEPIRIKQ--ENSGPPENDFPVYIVKQESDESRPONT 754
 QY 240 GSKG-----NKARSSSGEPKPLVR--AKGAQAGAPG-----GGEARLGQGSVPAP 282
 DB 755 NTPRSLTSLLLNSSQSSASEETVLRSDADSTGDQGLHQENSSNCKSE-WDSASQKSP 813
 QY 283 LALPSPQHLQKDEECAYCRDGGELICCGCGRAHPLALSLPLRITPGSTWRCCSCLQ 342
 DB 814 VAVGETRKEDDEDCAVQNGGELLCCCKCPKRVHLTCHVPLTLTFPSGEWICTCRD 873
 QY 343 ATVQEVQRAPEP 355
 DB 874 LSKPEVDYDCVP 886

RESULT 9
 CHD3_CAEEL STANDARD; PRT: 1787 AA.
 AC Q22516; Q18794;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Chromodomain helicase-DNA-binding protein 3 homolog (CHD-3).

GN CHD-3 OR T14G8.1.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdiltida; Rhabditoidea;
 CC Rhabdiltidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20530483; PubMed=11076750;
 RA von Zellewsky T., Palladino F., Brunschwig K., Tobler H., Hajnal A.,
 RA Mueller F.,
 RT "The C. elegans Mi-2 chromatin-remodelling proteins function in vulval
 RT cell fate determination."
 RL Development 127:5277-5284(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Matthews P., McMurray A.;
 RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
 CC - FUNCTION: Chromatin-remodelling protein that function in vulval
 CC cell fate determination.
 CC - SUBCELLULAR LOCATION: Nuclear (potential).
 CC - SIMILARITY: BELONGS TO THE SMF2/RAD54 HELICASE FAMILY.
 CC - SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
 CC - SIMILARITY: CONTAINS 2 CHROMO DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF308444; AAC29837.1; -
 DR EMBL: Z67884; CAA91810.1; -
 DR EMBL: Z67881; CAA91810.1; JOINED.
 DR EMBL: Z67881; CAA91798.1; -
 DR EMBL: Z67884; CAA91798.1; JOINED.
 DR Wormpep: T14G8.1; CE03657.
 DR Interpro: IPR000953; Chromo.
 DR Interpro: IPR001410; DEAD.
 DR Interpro: IPR002464; DEAD_box.
 DR Interpro: IPR001650; Helicase_C.
 DR Interpro: IPR000330; SMF2_N.
 DR Interpro: IPR001965; znf-PHD.
 DR Interpro: IPR001841; znf_ring.
 DR Pfam: PF00176; SMF2_N; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00385; chromo; 1.
 DR Pfam: PF00628; PHD; 2.
 DR SMART: SM00298; CHROMO; 2.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR SMART: SM00249; PHD; 2.
 DR SMART: SM00184; RING; 2.
 DR PROSITE: PS00598; CHROMO_1; FALSE_NEG.
 DR PROSITE: PS00513; CHROMO_2; 1.
 DR PROSITE: PS00690; DEAD_ATP_HELICASE; 1.
 DR PROSITE: PS01359; ZF-PHD; 1; 2.
 DR PROSITE: PS00016; ZF-PHD; 2; 2.
 KW Chromatin regulator; Nuclear protein; Repeat; Helicase; DNA-binding;
 KW ATP-binding; Zinc-finger.
 FT DOMAIN 59 62 POLY-LYS.
 FT ZN_FING 265 312 PHD-TYPE 1.
 FT ZN_FING 328 375 PHD-TYPE 2.
 FT DOMAIN 373 476 CHROMO 1.
 FT DOMAIN 501 583 CHROMO 2.
 FT DOMAIN 1287 1291 POLY-ARG.
 FT NP_BIND 641 648 ATP (POTENTIAL).
 FT SITE 763 766 DEAD BOX.
 FT SEQUENCE 1787 AA; 205254 MW; 1EFCE1FECE59740 CRC64;

Query Match

7.1%; Score 206; DB 1; Length 1787;


```

OY      169 VGCAGVAGGIIILQGVFEESGGKKICIQ-GEFPTYPKNEKDSGSGKNARSSGPKPLVR    257
        :|::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      590 LPVTGCGVAKGIHKKRKLQGLIIVKCIQTEDKGWFTPEIRIG--GHARKNW-RLSVR    645

OY      258 AKGAAGAAGGGEAR-LGQGCSVPAPLAL-----PSDPOLHKNDDECAV    301
        ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      646 C-----CGPRLMWLNENGLPPPRRRRYKKRIKLSQNNSVDPCM-RNLDECV    695

OY      302 CRDGGELICDCGCPRAFHLCISPLREIPSTGWRCSSG-----LATVGEVOPRAE    354
        |||||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB      696 CRDGELCCDCTGSRVFHEDECHIPPV-EARTPMWCICFRMKESPSSGCCCESEVLERQ    754

OY      355 PRPOE    359
        :|:
DB      755 MCPBE    759

RESULT 11
TF1G_HUMAN
AC   Q9UPN9; O95855; Q9UJ79; G9C017; PRT; 1127 AA.
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DI   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Transcription intermediary factor 1 gamma (TF1 gamma) (RET-fused gene
GN   7 protein) (Rfg protein) (Tripartite motif protein 33).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX   NCBI_TaxId=9606; [1]
RN   RP     MEDLINE FROM N.A. (ISOFORM ALPHA).
RX       MEDLINE=99144725; PubMed=10021217;
RA       Venturini L., You J., Stadler M., Gallen R., Lallemend V.,
RA       Koken M.H.M., Mattei M.-G., Ganter A., Chambon P., Losson R.,
RA       De The H.;
RT       "Tifgamma, a novel member of the transcriptional intermediary factor
RT       1 family.";
RN       Oncogene 18:1209-1217(1999).
RN       [2]
RP     SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).
RX       MEDLINE=21231161; PubMed=11331580;
RA       Raymond A., Meyron G., Fantozzi A., Merla G., Cairo S., Luzi L.,
RA       Ritgenelli D., Zanaria E., Messali S., Calnarca S., Guffanti A.,
RA       Minucci S., Pellicci P.G., Balibbio A.;
RT       "The tripartite motif family identifies cell compartments.";
RL       EMBO J. 20:2140-2151(2001).
RN       RN     SEQUENCE FROM N.A. (ISOFORM ALPHA).
RP     RC       TISSUE=Braint;.
RX       MEDLINE=99397452; PubMed=10470851;
RA       Kikuno R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,
RA       Tanaka A., Kotani H., Nomura N., Ohara O.;
RT       "Prediction of the coding sequences of unidentified human genes. XIV.
RT       The complete sequences of 100 new cDNA clones from brain which code
RT       for large proteins in vitro.";
RL       DNA Res. 6:197-205(1999).
RN       RN     [4]
RP     RP     SEQUENCE OF 76-1127 FROM N.A. (ISOFORM ALPHA).
RX       TISSUE=Thyroid;
RX       MEDLINE=99367802; PubMed=10439047;
RA       Klugbauer S., Rabes H.;
RT       "The transcription coactivator htfl1 and a related protein are fused
RT       to the ret receptor tyrosine kinase in childhood papillary thyroid
RT       carcinomas.";
RL       Oncogene 18:4388-4393(1999).
CC      CC      -1- FUNCTION: SEEMS TO ACT AS A TRANSCRIPTIONAL REPRESSOR.
CC      CC      -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC      CC      -1- ALTERNATIVE PRODUCTS: 2 isoforms: Alpha (shown here) and Beta;
CC      CC      are produced by alternative splicing.
CC      CC      -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC      CC      -1- SIMILARITY: CONTAINS 2 B BOX-TYPE ZINC FINERS.
```

```

CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: AF119043; AAD17259.1; -.
DR EMBL: AF220136; AAG53509.1; -.
DR EMBL: AF220137; AAG53510.1; -.
DR EMBL: AB029036; BA083065.1; ALT_INIT.
DR EMBL: AJ132948; CAB55313.1; -.
DR TRANSFAC: T04946; -.
DR Genem: HGNC:16290; TRIM33.
DR MIM: 605769; -.
DR InterPro: IPR003649; Bbox_C.
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR000315; ZnF_Box.
DR InterPro: IPR001965; ZnF_PHD.
DR InterPro: IPR001841; ZnF_Ring.
DR Pfam: PF00097; zf-C3HC4; 1.
DR Pfam: PF00439; bromodomain; 1.
DR Pfam: PF00628; PHD; 1.
DR Pfam: PF00643; zf-B_box; 2.
DR PRINTS: PRO1406; BBOXZNFINGER.
DR PRINTS: PRO0503; BROMODOMAIN.
DR SMART: SM00502; BBC; 1.
DR SMART: SM00336; BROX; 2.
DR SMART: SM00297; BROMO; 1.
DR SMART: SM00249; PHD; 2.
DR SMART: SM00184; RING; 2.
DR PROSITE: PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE: PS50014; BROMODOMAIN_2; 1.
DR PROSITE: PS50119; ZF_BBOX; 2.
DR PROSITE: PS01359; ZF_PHD_1; 1.
DR PROSITE: PS50016; ZF_PHD_2; 1.
DR PROSITE: PS00518; ZF_RING_1; 1.
DR PROSITE: PS50089; ZF_RING_2; 1.
DR Transcription regulation; Repressor; DNA-binding; Bromodomain;
DR Zinc-finger; Nuclear protein; Repeat; Coiled coil;
KW Alternative splicing.
FT DOMAIN 13 18 POLY-GLY.
FT ZN_FING 44 49 POLY-GLU.
FT ZN_FING 125 154 RING-TYPE.
FT ZN_FING 212 259 B_BOX-TYPE 1.
FT ZN_FING 271 312 B_BOX-TYPE 2.
FT DOMAIN 299 401 COILED COIL (POTENTIAL).
FT DOMAIN 545 550 POLY-THR.
FT ZN_FING 887 934 PHD-TYPE.
FT DOMAIN 974 1046 BROMODOMAIN.
FT VARSPIC 1041 1057 MISSING (IN ISOFORM BETA).
FT CONFLICT 89 V->E (IN REF. 4).
FT CONFLICT 451 453 PAA->LH (IN REF. 4).
FT CONFLICT 909 909 F->S (IN REF. 4).
FT CONFLICT 1037 1037 R->T (IN REF. 1).
FT CONFLICT 1114 1127 RKRKLSDEPRVPHIK->QKPKYK (IN REF. 1).
SQ SEQUENCE 1127 AA, 122521 MW, 02E380CFCDFDA3B CRC64;

Query Match 7.0%; Score 204; DB 1; Length 1127;
Best Local Similarity 25.9%; Pred. No. 0.00027;
Matches 68; Conservative 34; Mismatches 101; Indels 60; Gaps 10;
OY 115 PPAVKAIVPPRLPTKRRKASEBARAADAIPALTPRGTAAPSGSLAKAPPKKRESSAEQOR 174
Db 710 PQQPTSTNNPSP-----GPSALSP-----GSSGLSNHSTVYRPSTSTISGR 751
OY 175 LPLGIGIDITMSASVORAVAMSSGDV-----FGARAV---EGILIQVFEESGSKKCIQV 226
Db 752 GSCGSSGCT---AEKTSISFKSDQVKVKEGCTEDELCSFGGVQVQKTEIDGRRSACMLS 808

```

```

QY 227 GGE-----FYTPSKFEDSGSGKNKARS-----SSGPKPIYRKAQAQAAP 266
DB 809 SPESLTPPLSTNLHESLIDLASLENNHYKTEPADMNESSCKOSGLSLVNGKSPINSLM 868
QY 267 GGGEARLGGCGVPAPLALPSPDLQKNEDECACVCRDGGELICDCCPRAFHACLSP 326
DB 869 -HRSARIGGDGN-----NKDDP-----NEDMCATVCNGGDLCCCEKCPKPHLTCHPPT 917
QY 327 LREIPSGTWRCSSCLQATVQEVQ 349
DB 918 LLSFSGDWICTCFCRDICKREVE 940

RESULT 12
YA27_SCHPO
ID YA27_SCHPO STANDARD; PRT; 607 AA.
AC 009698;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C2E7.07c in chromosome I.
GN SPAC2E7.07c.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moesti D., Hilbert H.,
RA Borzym K., Zimmermann W., Medler H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Reinhardt R., Pohl T.M.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas R., Rochet M., Gallardin C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Paulsen I., Potashkin J.,
RA "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
RT -1- SIMILARITY: TO S.POMBE SPAC16C9.05 AND YEAST YMR075W.
CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z50142; CAA90494.1; -
DR InterPro: IPR001965; Znf_PHD.
DR Pfam: PF00628; PHD; 2.
DR SMART: SM00249; PHD; 2.

```

```

DR PROSITE: PS01359; ZF_PHD_1; 1.
DR PROSITE: PS50016; ZF_PHD_2; 2.
KM Hypothetical protein, Repeat, Zinc-finger.
FT ZN_FING 263 312 PHD-TYPE 1.
FT ZN_FING 406 459 PHD-TYPE 2.
SQ SEQUENCE 607 AA; 68776 MW; 43CB6520BA471AC CRC64;

Query Match
Best Local Similarity 22.7%; Pred. No. 0.0014;
Matches 83; Conservative 44; Mismatches 132; Indels 107; Gaps 18;

QY 151 TASPQSOLKAKPKPPRESSAEQRLPLGNGIOTMSAYORAVAMSSGDVPGARGAVEGIL 210
DB 169 TAKDLSDISSSMKKANNSK-----PLFSGLTFKANI-----PVPTSEVVTENNV 215
QY 211 IQQYFESGSKKCIQVGEFFTPSKFEDSGSGKNKARSSSGPKPIYRKAQAQAAPGGE 270
DB 216 TRNVTVYSNOK-----HIGNE-----SENFDM-----EGRAEDISSNEL- 251
QY 271 ARLGQGGVAPAPLALPSPDLQKNEDECACVCRDGGELICDCCPRAFHACLSPLE- 329
DB 252 -----FPEEYF-----YRYNDYCSACHGPNFLCCTCPSHFPTCIDPIEEK 297
QY 330 -IPSGTWRCSSCLQATV-----QEVQPRAEPR-----POEPPVETPLDP 368
DB 298 NLPDDAWYCNCKHSLYLNLEDEQELSNVKEEGTVDVMQLCTYIDSHNPQFHLPH 357
QY 369 GL-----KSAGEEVGPRGPEPLAGMDTLYVKNLPAP-----PSAAPPLDLSALHPLC 419
DB 358 SISSFFRGVSGVGEYIE-----TDVLRKLSRSRNGSERDPL-LKSSGTPILC 409
QY 420 VGEQGQNLAPGARGVCG-DGTDLVLRCTHCAAFHWRCHFPAGTSRPTGLR---CRSC 475
DB 410 F-----RCHSALVQSILADICYSVHHPDCLNPLPLTP-SNLKMKMCPNH 456
QY 476 SGDVTP 481
DB 457 SDHVT 462

RESULT 13
YA25_SCHPO
ID YA25_SCHPO STANDARD; PRT; 404 AA.
AC 009819;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C16C9.05 in chromosome I.
GN SPAC16C9.05.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moesti D., Hilbert H.,

```

RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Gerbulet L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shipakowski G.V., Ussery D., Barrell B.G., Nurse P.,
 RA "The genome sequence of Schistosoma mansoni pombae";
 RL Nature 415:871-880(2002).
 CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: 254366; CAA91193.1;
 DR InterPro: IPR001965; Znf_PHD.
 DR Pfam: PF00628; PHD_1.
 DR SMART: SM00249; PHD_1.
 DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS50016; ZF_PHD_2; 1.
 DR KW Hypothetical protein; Zinc-finger.
 FT ZN_FING 117 166 PHD-TYPE.
 FT SEQUENCE 404 AA; 44984 MW; CB9412E792C1C5D4 CRC64;
 SO
 Query Match
 Best Local Similarity 6.2%; Score 179.5; DB 1; Length 404;
 Matches 67; Conservative 27; Mismatches 94; Indels 75; Gaps 12;
 QY 234 KNEDECAVCRDGGELICCGCPAPHLACLSPRL--REIPSGTWKSSCCQATVQEVQPR 351
 DB 115 RNVDCSACGGGRGRLFCGCGCFCSFHLSCLEPLTPENIPREGSMFCVTC--SIKSHNP 170
 QY 352 AEERPOEP-----PVETPLPGRLRSAGEEV--RPPG---EPLAGMDTTL 392
 DB 171 ----PKHPLSLWSQLYDWIDQNSQYRLPDDLVHYHGISGDMGKAYKETGEEDT-- 223
 QY 393 YKHLPAAPSAAPRLGDLSSALHPLLCVGRPEGOQLNAPARCGVCGDGT--VLKCTH 448
 DB 224 --DEFSALPTGSSITNL-----AYCGYCKSPMGACVWYGCOL 259
 QY 449 CAAAFHWRCHPAGTSRPGTGLKRCSGGDVTPAPVEGVLASPARLAPGAKDPTASH- 507
 DB 260 CDTEYHNKCKEHA-----KKCSHD--SIKKGRVPEKMAVIRPLVLDITSTW 306
 QY 508 -EPALHNDLESLSEHTFDGIL 529
 DB 307 LNPVKWISGQFLMGEPSSDEL 329
 RESULT 14
 YHLL_EBV STANDARD; PRT; 660 AA.
 AC P03181;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DE 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Hypothetical BHLF1 protein.
 OS Epstein-Barr virus (strain 95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OK NCBI_TaxID=10377;
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,

RA Tuftnell P.S., Barrell B.G.;
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome";
 RL Nature 310:207-211(1984).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: A01555; -; NOT_ANNOTATED_CDS.
 DR PIR: A03742; Q0BE3.
 KW Hypothetical protein; Early protein; Repeat.
 FT DOMAIN 149 648 4 x 125 AA TANDDEM REPEATS.
 FT REPEAT 149 273 1.
 FT REPEAT 274 398 2.
 FT REPEAT 399 523 3.
 FT REPEAT 524 648 4.
 SO SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;
 Query Match
 Best Local Similarity 5.8%; Score 168.5; DB 1; Length 660;
 Matches 120; Conservative 26; Mismatches 192; Indels 175; Gaps 25;
 QY 109 PKGGRK-----PAVRLVPRPLPKRKASEEAAARAALPR-----GTASPGS 156
 DB 172 PGAGRGPSGPTGGRRAARQA-----FETPA--APGCGGAABVSGATPHERGSGPADPPA 225
 QY 157 OLKAKP-----PKPESSAEQRLPLGNGIOTMSASVAVRAMSSGDVPCAR----- 203
 DB 226 AARLPREKQERPLPDLLAAORCPAGRPPTTSGAAQRTNRRRPGCRRAANRPGCRRTR 285
 QY 204 ---GAVEGILLQVRESGSKKCIQVGEFTYPSKFEDSGSKNKARSSGKRLVRAKG 260
 DB 286 RRSAGRG-----HPP-----PGAQ--RPSGPTGGR--AARG 315
 QY 261 AAG--AARGGEARLGGQGVAPRL--ALPSPDLQKNEDCAVCRDGGELICDGCPR 316
 DB 316 APTRAAPRGFGGAAVRGATPHERGSGPADPPAARLPREKQERPLPDLLAAORCPA 375
 QY 317 AFHILACSLPLR-----ETPSG-----TWRCSSCIQATVQEVQPRAEPR 355
 DB 376 G-----PPTRSGAAARQTHRRRPPGCRSARNPGCPRTWRRRSGAQ----- 416
 QY 356 RPOEPVETPLPRLRSAGEEVRRPGRPLAGMDTTLVYKHLPAAPSAAPRLGDLSSALH 415
 DB 417 RGHPRPGAGRPSGPTGGRPAARFAGPTPA-----PGGGGAAYP--SGATPH 463
 QY 416 PLLCVG-----PEGO-----QNLAPGARC-----GYVCGDGT-- 442
 DB 464 PERGSGPADPPAARLPREKQERPLPDLLAAQCRPAGRPPTTSGAAQRTNRRRPGCPR 523
 QY 443 VLKCTHCAAFHWRCHPAGTSRPGTGLKRCSCG-----DVTPAPVEGVLAPS 491
 DB 524 SARNPCCRTTWRRRSAGQRCNRPFGAGORPSGPTGGRPAARFAGPTPAARPGGGAAPVS 583
 QY 492 PARLAP-----GPAKDDTAS-----HEPALHRD 514
 DB 584 GATPHERGSGPADPPAARLPREKQERPLPD 616
 RESULT 15
 YMW5_YEAST STANDARD; PRT; 684 AA.
 AC Q04779;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical 78.8 kDa protein in ABF2-CHL12 intergenic region.
 GN YMR075W OR YW9916.14.
 OS Saccharomyces cerevisiae (Baker's yeast).

CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO S.POMBE SPAC16C9.05 AND SPAC2F7.07C.
CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC
DR EMBL; 248952; CA88800.1; -
DR SGD; S0004680; YMR075W.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW Hypothetical protein; Zinc-finger.
FT ZN_FING 260 309 PHD-TYPE.
FT DOMAIN 15 30 POLY-SER.
SQ SEQUENCE 684 AA: 78836 MW: CRC282DC9A782E5C CRC64;

Query Match 5.5%; Score 161; DB 1; Length 684;
Best Local Similarity 21.7%; Pred. No. 0.038;
Matches 48; Conservative 34; Mismatches 85; Indels 54; Gaps 8;

QY 294 KNEDECAVCRDGGELICDGGPRAFLACLSPRL--REIPSGTWRCSSC-----LQA 343
DB 258 ENEDFCSACNOSGFELCDTCPSKHFHCLDPPIDPNNILPKGDWHCNECKFKIFINNSMA 317
QY 344 TVQEVQPRAEPR-----POEPVETPLPPGLRSAGEVR--GPGE----- 383
DB 318 TLKKIESNFIKONNNVAKIFAKLLFNIDSHNPKQFOLPNYIKETFPRAVKTSRGQYSDEND 377
QY 384 --PLAG--MDTLYYKHLPAAPSAAPLPGLDSSALHPLLCVGPPEGQONLAPGARCGVCGD 439
DB 378 KIPLFTDROLFNTSYGOSITKLDSTYNDPTFHIDNSGKFLICY-----KCNQTRL 425
QY 440 GT-----DVLRCTHCAAFHWRCHFRPAGTSRPGTGLRC 472
DB 426 GSWSHPENSRLLIMTCDCYQTFPHLDYCPRASFKNLGSKWK 466

Search completed: March 13, 2003, 17:53:56
Job time : 20 secs

